

# Diffusion Adaptation over Multi-Agent Networks with Wireless Link Impairments

Reza Abdolee, *Student Member, IEEE*, Benoit Champagne, *Senior Member, IEEE*,  
and Ali H. Sayed, *Fellow, IEEE*

**Abstract**—We study the performance of diffusion least-mean-square algorithms for distributed parameter estimation in multi-agent networks when nodes exchange information over wireless communication links. Wireless channel impairments, such as fading and path-loss, adversely affect the exchanged data and cause instability and performance degradation if left unattended. To mitigate these effects, we incorporate equalization coefficients into the diffusion combination step and update the combination weights dynamically in the face of randomly changing neighborhoods due to fading conditions. When channel state information (CSI) is unavailable, we determine the equalization factors from pilot-aided channel coefficient estimates. The analysis reveals that by properly monitoring the CSI over the network and choosing sufficiently small adaptation step-sizes, the diffusion strategies are able to deliver satisfactory performance in the presence of fading and path loss.

**Index Terms**—Distributed estimation, diffusion LMS, link-failure, fading channels, wireless sensor networks, combination policy.

## 1 INTRODUCTION

**D**IFFUSION least-mean squares (LMS) algorithms can serve as efficient and powerful mechanisms for solving distributed estimation and optimization problems over networks in real-time, in response to streaming data originating from different locations [1]–[5]. Owing to their decentralized processing structure, simplicity of implementation, and adaptive learning capabilities, these algorithms are particularly well-suited for applications involving multi-agent wireless networks, where energy and radio resources are generally limited [?], [8], [9]. Consensus strategies can also be used for distributed estimation purposes [10]–[15]. However, it was shown in [16] that for constant step-size adaptation, network states can grow unbounded due to an inherent asymmetry in the consensus dynamics. The same problem does not occur for diffusion strategies, and for this reason, we focus on these algorithms in this work.

Diffusion strategies have been widely investigated in networks with static topologies in which the communication links between agents remain invariant with respect to time [2], [7], [17]–[22]. Under such conditions, these strategies converge in the mean and mean-square error sense in the slow adaptation regime [2], [3], [5], [16], [23]. Previous studies have also examined the effect of noisy communication links on the performance of these algorithms on network with static topologies [24]–

[27]. The main conclusion drawn from these works is that performance degradation occurs unless the combination weights used at each node are adjusted to counter the effect of noise.

The static link topology assumption, however, is restrictive in applications in wireless communications and sensor network systems. For example, in mobile networks where the agents are allowed to change their position over time, the signal-to-noise ratio (SNR) over the communication links between nodes will vary due to the various channel impairments, including path loss, multi-path fading and shadowing. Consequently, the set of nodes with which each agent can communicate (called neighborhood set) will also change over time, as determined by the link SNR, and the network topology is therefore intrinsically dynamic. It is therefore essential to study the performance of diffusion strategies over networks with time-varying (dynamic) topology and characterize the effects of link activity (especially link failure) on their convergence and stability.

The problem of link imperfection was also investigated in other classes of distributed algorithms, such as consensus [28]–[32] and subgradient algorithms [9], [33]. In [28], [29] and [33], the authors have examined the performance of consensus algorithms over networks with link failures, where links are established according to some predefined probabilities. They assumed that once a link is activated at a given iteration the data received through it will be undistorted. References [31], [32] have taken into account the effects of link and quantization noise in addition to link failure and investigated the network convergence and stability. A more realistic network scenario was considered in [30], [34] where the probabilities of link failure are obtained using a fading channel model and SNR of the received signals. However, the data received from a neighboring node is

- R. Abdolee and B. Champagne are with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, H3A 0E9 Canada (e-mail: reza.abdolee@mail.mcgill.ca, benoit.champagne@mcgill.ca). The work of B. Champagne and R. Abdolee was supported in part by an NSERC Discovery grant.
- A. H. Sayed is with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA (e-mail: sayed@ee.ucla.edu). The work of A. H. Sayed was supported in part by NSF CCF-1011918 and ECCS-1407712.

assumed to be error-free when the corresponding link is active.

In this paper, we study the performance of diffusion estimation strategies over networks with time-varying topologies where the information exchange between agents occurs over noisy wireless links that are also subject to fading and path loss<sup>1</sup>. Our contributions are as follows. We extend the application of diffusion LMS strategies from multi-agent networks with ideal communication links to sensor networks with fading wireless channels. Under fading and path loss conditions over wireless links, the neighborhood sets become dynamic, with nodes leaving or entering neighborhoods depending on the quality of the links as defined by the instantaneous SNR conditions. Our analysis will show that if each node knows the channel state information (CSI) of its neighbors, the effects of fading and path-loss can be mitigated by incorporating local equalization coefficients into the diffusion updates. When CSI is not available to the nodes, we explain how the equalization coefficients can be evaluated from a pilot-assisted estimation process along with the main parameter estimation task of the network. We also examine the effect of channel estimation errors on the performance and convergence of the modified algorithms in terms of a mean-square-error metric. We establish conditions under which the network is mean-square stable for both known and unknown CSI cases. The analysis reveal that when CSI is known, the modified diffusion algorithms are asymptotically unbiased and converge in the slow adaptation regime. In contrast, the parameter estimates will become biased when the CSI are obtained through pilot-aided channel estimation. Nevertheless, the size of the bias can be made small by increasing the number of pilot symbols or increasing the link SNR.

The paper is organized as follows. In Section 2, we explain the network signal model. In Section 3, we review the standard diffusion strategies and introduce a modification for distributed estimation over wireless networks. We analyze the convergence and stability of the proposed algorithms in Section 4. We present the simulation results in Section 5, and conclude the paper in Section 6.

*Notation:* Matrices are represented by upper-case and vectors by lower-case letters. Boldface fonts are reserved for random variables and normal fonts are used for deterministic quantities. Superscript  $(\cdot)^T$  denotes transposition for real-valued vectors and matrices while  $(\cdot)^*$  denotes conjugate transposition for complex-valued vectors and matrices. The symbol  $\mathbb{E}[\cdot]$  is the expectation operator,  $\text{Tr}(\cdot)$  represents the trace of its matrix argument and  $\text{diag}\{\cdot\}$  extracts the diagonal entries of a matrix, or constructs a (block) diagonal matrix using its argument. A set of vectors are stacked into a column vector by  $\text{col}\{\cdot\}$ . The  $\text{vec}(\cdot)$  operator vectorizes a matrix by

stacking its columns on top of each other and  $\text{bvec}(\cdot)$  is the block-vectorization operator [1]. The symbol  $\otimes$  denotes the standard Kronecker product, and the symbol  $\otimes_b$  represents the *block Kronecker product* [1].

## 2 NETWORK SIGNAL MODEL

Consider a set of  $N$  sensor nodes that are distributed over a geographical area. At time instant  $i \in \{0, 1, \dots\}$ , each node  $k \in \{1, 2, \dots, N\}$  collects data  $\mathbf{d}_k(i)$  and  $\mathbf{u}_{k,i}$  that are related to an unknown parameter vector  $\mathbf{w}^o \in \mathbb{C}^{M \times 1}$  via the following relation:

$$\mathbf{d}_k(i) = \mathbf{u}_{k,i} \mathbf{w}^o + \mathbf{v}_k(i) \quad (1)$$

where  $\mathbf{d}_k(i) \in \mathbb{C}$ ,  $\mathbf{u}_{k,i} \in \mathbb{C}^{1 \times M}$  and  $\mathbf{v}_k(i) \in \mathbb{C}$  are, respectively, the scalar measurement, the node's regression vector and the measurement noise.

**Assumption 1.** *The variables in the linear regression model (1) satisfies the following conditions:*

- a) *The regression vectors  $\{\mathbf{u}_{k,i}\}$  are zero-mean, i.i.d. in time, and independent over space, with covariance matrices  $R_{u,k} = \mathbb{E}[\mathbf{u}_{k,i}^* \mathbf{u}_{k,i}] > 0$ .*
- b) *The measurement noise  $\{\mathbf{v}_k(i)\}$  are zero-mean, i.i.d. in time, and independent over space, with variances  $\sigma_{v,k}^2$ .*
- c) *The regression vectors  $\mathbf{u}_{k_1,i_1}$  and the noise  $\mathbf{v}_{k_2,i_2}$  are mutually independent for all  $k_1, k_2, i_1$  and  $i_2$ .*

Node  $\ell$  is said to be a neighbor of node  $k$  if its distance from node  $k$  is less than a preset transmission range  $r_o$  [36], which for simplicity is assumed to remain constant over the given geographical area. The set of all neighbors of node  $k$ , including node  $k$  itself, is denoted by  $\mathcal{N}_k$ . Nodes are allowed to communicate with their neighbors only, but due to channel impairments, certain links may fail. Hence, at any given time  $i$ , only a subset of the nodes in  $\mathcal{N}_k$  can communicate with node  $k$ .

The objective of the network is to estimate the unknown parameter vector  $\mathbf{w}^o$  in a distributed manner when the data exchange between the agents occurs over noisy wireless links that are also subject to fading and path loss. In particular, we assume that the transmit signal  $\psi_{\ell,i} \in \mathbb{C}^{M \times 1}$  from node  $\ell \in \mathcal{N}_k \setminus \{k\}$  to node  $k$  at time  $i$  experiences channel distortion of the following form (see Fig. 1):

$$\psi_{\ell k,i} = \mathbf{h}_{\ell,k}(i) \sqrt{\frac{P_t}{r_{\ell,k}^\alpha}} \psi_{\ell,i} + \mathbf{v}_{\ell k,i}^{(\psi)} \quad (2)$$

where  $\psi_{\ell k,i} \in \mathbb{C}^{M \times 1}$  is the distorted estimate received by node  $k$ ,  $\mathbf{h}_{\ell,k}(i) \in \mathbb{C}$  denotes the fading channel coefficient over the wireless link between nodes  $k$  and  $\ell$ ,  $P_t \in \mathbb{R}^+$  is the transmit signal power,  $r_{\ell,k} = r_{k,\ell} \in \mathbb{R}^+$  is the distance between nodes  $\ell$  and  $k$ ,  $\alpha \in \mathbb{R}^+$  is the path loss exponent and  $\mathbf{v}_{\ell k,i}^{(\psi)} \in \mathbb{C}^{M \times 1}$  is the additive noise vector with covariance matrix  $\sigma_{v,\ell k}^{2(\psi)} I_M$ . We define  $\psi_{kk,i} \triangleq \psi_{k,i}$  to maintain consistency in the notation.

**Assumption 2.** *The fading channel coefficients and the link noise in (2) satisfy the following conditions:*

1. A short preliminary version of this work was presented in the IEEE International Conference on Communication (ICC), June 2013 [35].

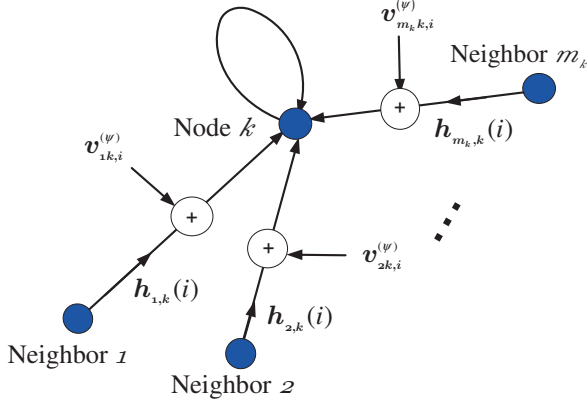


Fig. 1: Node  $k$  receives distorted data from its  $m_k = |\mathcal{N}_k|$  neighbors at time  $i$ . The data are affected by channel fading coefficients,  $\mathbf{h}_{\ell,k}(i)$ , and communication noise  $\mathbf{v}_{\ell,k,i}^{(\psi)}$ .

- a) The time-varying channel coefficients  $\mathbf{h}_{\ell,k}(i)$  follow the Clark's model [37], i.e., they are independent circular Gaussian random variables with zero mean and variance  $\sigma_{h,\ell,k}^2$ .
- b)  $\{\mathbf{h}_{\ell,k}(i)\}$  are independent over space and i.i.d. over time.
- c) The noise vectors  $\{\mathbf{v}_{\ell,k,i}^{(\psi)}\}$  are zero-mean, i.i.d. in time and independent over space.
- d) The channel coefficients,  $\mathbf{h}_{\ell,k_1}(i_1)$ , the noise vectors,  $\mathbf{v}_{\ell k_2,i_2}^{(\psi)}$ , the regression vectors,  $\mathbf{u}_{k_3,i_3}$  and the measurement noise,  $\mathbf{v}_{k_4,i_4}$ , are mutually independent for all  $k_j$  and  $i_j$  with  $j \in \{1, 2, 3, 4\}$ .

It is also assumed that nodes are aware of the positions of their neighbors through some positioning techniques and, therefore,  $r_{\ell,k}$ ,  $\ell \in \mathcal{N}_k$  is known to node  $k$ . A transmission from node  $\ell$  to node  $k$  at time  $i$  is said to be successful if the SNR between nodes  $\ell$  and  $k$ , denoted by  $\varsigma_{\ell,k}(i)$ , exceeds some threshold level  $\varsigma_{\ell,k}^o$ . The threshold level is defined as the SNR in the non-fading link scenario and is computed as:

$$\varsigma_{\ell,k}^o \triangleq \frac{P_t}{\sigma_{v,\ell,k}^{2(\psi)} r_{\ell,k}^\alpha} \quad (3)$$

In fading conditions, the instantaneous SNR is:

$$\varsigma_{\ell,k}(i) = \frac{|\mathbf{h}_{\ell,k}(i)|^2 P_t}{\sigma_{v,\ell,k}^{2(\psi)} r_{\ell,k}^\alpha} \quad (4)$$

When transmission is successful, we have  $\varsigma_{\ell,k}(i) \geq \varsigma_{\ell,k}^o$  which amounts to the condition:

$$|\mathbf{h}_{\ell,k}(i)|^2 \geq \nu_{\ell,k} \quad (5)$$

where  $\nu_{\ell,k} = \left(\frac{r_{\ell,k}}{r_o}\right)^\alpha$ . Since  $\mathbf{h}_{\ell,k}(i)$  has a circular complex Gaussian distribution, the squared magnitude  $|\mathbf{h}_{\ell,k}(i)|^2$  is exponentially distributed with parameter  $\lambda'_{\ell,k} = 1/\sigma_{h,\ell,k}^2$  [38]. Considering this fact, the probability of successful transmission is then given by:

$$p_{\ell,k} = \Pr(|\mathbf{h}_{\ell,k}(i)|^2 \geq \nu_{\ell,k}) = e^{-\lambda'_{\ell,k} \nu_{\ell,k}} \quad (6)$$

This expression shows that the probability of successful transmission decreases as the distance between two nodes increases. As such, the link between neighboring nodes is not guaranteed to be connected all the time, implying that the network topology is time-varying. Under this condition, we redefine the neighborhood set of node  $k$  as a time-varying set consisting of all nodes  $\ell \in \mathcal{N}_k$  for which  $\varsigma_{\ell,k}(i)$  exceeds  $\varsigma_{\ell,k}^o$  provided that node  $k$  knows the CSI of nodes  $\ell \in \mathcal{N}_k$ . In this way, the effective neighborhood set of each node  $k$  becomes random and we, therefore, denote it by  $\mathcal{N}_{k,i}$ . This implies that  $\mathcal{N}_{k,i} \subset \mathcal{N}_k$  for all  $i$ .

### 3 DISTRIBUTED ESTIMATION OVER WIRELESS CHANNELS

We first briefly review the standard diffusion LMS strategies for estimation of  $w^o$  over multi-agent networks with ideal links. We then elaborate on how to modify these strategies to enable the estimation of  $w^o$  in the presence of fading and wireless channel impairments.

#### 3.1 Diffusion Strategies over Ideal Communication Channels

In the context of mean-square-error estimation, diffusion strategies are stochastic gradient algorithms that can be used for the distributed minimization of the following global objective function [2], [3]:

$$J^{\text{glob}}(w) = \sum_{k=1}^N \mathbb{E} |d_k(i) - \mathbf{u}_{k,i} w|^2 \quad (7)$$

There are various forms of diffusion depending on the order in which the relevant adaptation and combination steps are performed. The so-called Adapt-then-Combine (ATC) strategy takes the following form:

$$\psi_{k,i} = \mathbf{w}_{k,i-1} + \mu_k \mathbf{u}_{k,i}^* [d_k(i) - \mathbf{u}_{k,i} \mathbf{w}_{k,i-1}] \quad (8)$$

$$\mathbf{w}_{k,i} = \sum_{\ell \in \mathcal{N}_k} a_{\ell,k} \psi_{\ell,i} \quad (9)$$

where  $\mu_k > 0$  is the step-size used by node  $k$ , and the  $a_{\ell,k}$  denote nonnegative entries of a left-stochastic matrix  $A$  that satisfy:

$$a_{\ell,k} = 0 \text{ if } \ell \notin \mathcal{N}_k \text{ and } \sum_{\ell \in \mathcal{N}_k} a_{\ell,k} = 1 \quad (10)$$

In this implementation, (8) is an adaptation step where node  $k$  updates its intermediate estimate  $\mathbf{w}_{k,i-1}$  to  $\psi_{k,i}$  using its measured data  $\{\mathbf{u}_{k,i}, d_k(i)\}$ . Then (9) is a combination step in which each node  $k$  combines its intermediate estimate  $\psi_{k,i}$  with that of its neighbors to obtain  $\mathbf{w}_{k,i}$ .

While the above algorithm works well over ideal communication channels, some degradation occurs when the exchange of information between neighboring nodes is subject to noise, as explained in [?], [24]–[26], [39], [40]. In this work, we move beyond these earlier studies and

examine the performance of diffusion strategies over fading wireless channels. We also suggest modifications to the update equations to counter the effect of fading.

### 3.2 Diffusion Strategies over Wireless Channels

We are initially motivated to replace the combination step in (9) by

$$\mathbf{w}_{k,i} = \sum_{\ell \in \mathcal{N}_k} a_{\ell,k} \bar{\psi}_{\ell,i} \quad (11)$$

where  $\bar{\psi}_{\ell,i}$  is a refined version of the distorted estimate  $\psi_{\ell k,i}$  that node  $k$  receives. The refinement is computed through a scaling equalization step of the form:

$$\bar{\psi}_{\ell,i} = g_{\ell,k}(i) \psi_{\ell k,i} \quad (12)$$

where the scalar gain  $g_{\ell,k}(i)$  is an equalization coefficient to be chosen to counter the effect of fading. Recall that  $\psi_{\ell k,i}$  is related to  $\psi_{\ell,i}$  via (2). Moreover, since each node  $k$  uses data from nodes  $\ell \in \mathcal{N}_k$  whose instantaneous SNR,  $\varsigma_{\ell k}(i)$ , exceeds the threshold  $\varsigma_{\ell k}^o$ , then we need to further adjust (9) and replace  $\mathcal{N}_k$  and  $a_{\ell,k}$ , respectively, with  $\mathcal{N}_{k,i}$  and  $a_{\ell,k}(i)$ . This leads to:

$$\mathbf{w}_{k,i} = \sum_{\ell \in \mathcal{N}_{k,i}} a_{\ell,k}(i) g_{\ell,k}(i) \psi_{\ell k,i} \quad (13)$$

Therefore, in wireless sensor networks, the ATC diffusion strategy takes the form presented in Algorithm 1.

---

#### Algorithm 1 : Diffusion ATC over Wireless Channels

---

$$\psi_{k,i} = \mathbf{w}_{k,i-1} + \mu_k \mathbf{u}_{k,i}^* [\mathbf{d}_k(i) - \mathbf{u}_{k,i} \mathbf{w}_{k,i-1}] \quad (14)$$

$$\mathbf{w}_{k,i} = \sum_{\ell \in \mathcal{N}_{k,i}} a_{\ell,k}(i) g_{\ell,k}(i) \psi_{\ell k,i} \quad (15)$$


---

One way to compute the equalization coefficients in (42) is to employ the following zero-forcing type construction:

$$g_{\ell,k}(i) = \begin{cases} \frac{\mathbf{h}_{\ell,k}^*(i)}{|\mathbf{h}_{\ell,k}(i)|^2} \sqrt{\frac{r_{\ell,k}^\alpha}{P_t}} & \text{if } \ell \in \mathcal{N}_{k,i} \setminus \{k\} \\ 1 & \text{if } \ell = k \end{cases} \quad (16)$$

Alternatively, if the noise variances  $\sigma_{v,\ell k}^{2(\psi)}$  are known, then one could also use minimum mean-square-error (MMSE) estimation to obtain the equalization coefficients. For simplicity, we continue with (16). By switching the order of the adaption and combination steps in Algorithm 1, we will obtain the Combine-then-Adapt (CTA) diffusion strategy, which is presented below as Algorithm 2. In (17),  $\mathbf{w}_{\ell k,i}$  is the estimate of the global parameter at node  $\ell$  that undergoes similar path loss, fading and noise as  $\psi_{\ell k,i}$  described by (2).

The combination coefficients  $a_{\ell,k}(i)$  in (13) now become random and time-dependent because the neighborhood

---

#### Algorithm 2 : Diffusion CTA over Wireless Channels

---

$$\psi_{k,i-1} = \sum_{\ell \in \mathcal{N}_{k,i}} a_{\ell,k}(i) g_{\ell,k}(i) \mathbf{w}_{\ell k,i-1} \quad (17)$$

$$\mathbf{w}_{k,i} = \psi_{k,i-1} + \mu_k \mathbf{u}_{k,i}^* [\mathbf{d}_k(i) - \mathbf{u}_{k,i} \psi_{k,i-1}] \quad (18)$$


---

sets,  $\mathcal{N}_{k,i}$ , are also evolving with time. Moreover, they need to satisfy

$$a_{\ell,k}(i) = 0 \text{ if } \ell \notin \mathcal{N}_{k,i} \text{ and } \sum_{\ell \in \mathcal{N}_{k,i}} a_{\ell,k}(i) = 1 \quad (19)$$

The randomness of  $a_{\ell,k}(i)$  can be further clarified by resorting to (5). The communication between nodes  $\ell$  and  $k$  is successful if (5) is satisfied; otherwise, the link between them fails. When the link fails, the associated combination weight  $a_{\ell,k}(i)$  must be set to zero, which in turn implies that other combination coefficients of node  $k$  need to be adjusted to satisfy (19). This suggests that the neighborhood set  $\mathcal{N}_{k,i}$  has to be updated whenever one of the neighborhood link SNR crosses the threshold in either direction:

$$\mathcal{N}_{k,i} = \left\{ \ell \in \mathcal{N}_k \mid \varsigma_{\ell k}(i) \geq \varsigma_{\ell k}^o \right\} \quad (20)$$

In practice, since  $\varsigma_{\ell k}(i)$  may not be measurable, we use (3)-(4) and (5) to update the neighborhood set as:

$$\mathcal{N}_{k,i} = \left\{ \ell \in \mathcal{N}_k \mid |\mathbf{h}_{\ell k}(i)|^2 \geq \nu_{\ell k} \right\} \quad (21)$$

Motivated by these considerations, we propose the following dynamic structure to adjust the combination weights over time:

$$a_{\ell,k}(i) = \begin{cases} \gamma_{\ell,k} \mathcal{I}_{\ell,k}(i), & \text{if } \ell \in \mathcal{N}_{k,i} \setminus \{k\} \\ 1 - \sum_{\ell \in \mathcal{N}_{k,i} \setminus \{k\}} a_{\ell,k}(i), & \text{if } \ell = k \end{cases} \quad (22)$$

where the  $\gamma_{\ell,k}$  are fixed, positive combination weights that node  $k$  assigns to its neighbors  $\ell \in \mathcal{N}_{k,i}$ . To ensure  $a_{k,k}(i) > 0$ , these weights need to satisfy:

$$\sum_{\ell \in \mathcal{N}_{k,i} \setminus \{k\}} \gamma_{\ell,k} < 1 \quad (23)$$

It can be verified that if each node  $k$  obtains the coefficients  $\gamma_{\ell,k}$  for the time-invariant neighborhood set  $\mathcal{N}_k$  according to well-known left or doubly-stochastic matrix combination rules (e.g., uniform averaging rule or Metropolis rule) then the condition (23) will be satisfied. In (22), the quantity  $\mathcal{I}_{\ell,k}(i)$  is defined as:

$$\mathcal{I}_{\ell,k}(i) = \begin{cases} 1, & \text{if } \ell \in \mathcal{N}_{k,i} \\ 0, & \text{otherwise} \end{cases} \quad (24)$$

When transmission from node  $\ell$  to node  $k$  is successful  $\mathcal{I}_{\ell,k}(i) = 1$ , otherwise,  $\mathcal{I}_{\ell,k}(i) = 0$ . In this way, the entries  $a_{\ell,k}(i)$  satisfy condition (19). From (20) and (24), we see that the indicator operator,  $\mathcal{I}_{\ell,k}(i)$ , is a random variable with Bernoulli distribution for which the probability of success,  $p_{\ell,k}$ , is given by the exponential function (6).



### 3.3 Modeling the Impact of Channel Estimation Errors

In Algorithms 1 and 2, it is assumed that each node  $k$  knows the channel fading coefficients  $\mathbf{h}_{\ell,k}(i)$ , which are needed in (16). In practice, this information is usually recovered by means of an estimation step. Consequently, some additional estimation errors will be introduced into the network.

There are many ways by which the fading coefficients can be estimated. For example, we may assume that the transmitted data from node  $\ell$  to node  $k$  carries two data types, namely, pilot symbols (training data) denoted by  $\mathbf{s}_\ell(i)$ , and data symbols  $\boldsymbol{\psi}_{\ell,i}$  or  $\mathbf{w}_{\ell,i-1}$ . The training data are used for channel estimation and the data symbols are the intermediate estimates of the unknown parameter vector,  $\mathbf{w}^o$ , which are used to update the network estimate at node  $k$ . According to (2), the received training data at node  $k$  and time  $i$  is affected by fading and noise, i.e.,

$$\mathbf{y}_{\ell,k}(i) = \mathbf{h}_{\ell,k}(i) \sqrt{\frac{P_t}{r_{\ell,k}^\alpha}} \mathbf{s}_\ell(i) + \mathbf{v}_{\ell,k}^{(y)}(i) \quad (25)$$

where  $\mathbf{v}_{\ell,k}^{(y)}(i)$  is a zero-mean additive white Gaussian noise with variance  $\sigma_{v,\ell,k}^{(y)2}$ . It is reasonable to assume that  $\sigma_{v,\ell,k}^{(y)2} = \sigma_{v,\ell,k}^{(\psi)2}$ . The number of training symbols used depends on the specific application requirements and the time scale variations of the channel. If we use a single training data to estimate each coefficient and assume that nodes  $k \in \{1, 2, \dots, N\}$  sends  $\mathbf{s}_k(i) = 1$  as training symbols, the least-squares estimation method gives the following estimate:

$$\hat{\mathbf{h}}_{\ell,k}(i) = \sqrt{\frac{r_{\ell,k}^\alpha}{P_t}} \mathbf{y}_{\ell,k}(i) \quad (26)$$

**Remark 1.** If we use an alternative way to find the threshold SNR,  $\nu_{\ell,k}^o$  without using distance information, then (25) can be expressed as  $\mathbf{y}_{\ell,k}(i) = \beta_{\ell,k}(i) \mathbf{s}_\ell(i) + \mathbf{v}_{\ell,k}^{(y)}(i)$ , where  $\beta_{\ell,k}(i) = \mathbf{h}_{\ell,k}(i) (P_t / r_{\ell,k}^\alpha)^{1/2}$ . In this form the fading coefficient and path loss are combined into a new channel coefficient  $\beta_{\ell,k}(i)$  that implicitly includes the distance information. In this case, to estimate the channel coefficients,  $\beta_{\ell,k}(i)$ , unlike (26), the distance information are not required.

From (25), it can be seen that  $\mathbf{y}_{\ell,k}(i)$  is composed of the sum of two independent circular Gaussian random variables. It follows that  $\mathbf{y}_{\ell,k}(i)$  will have circular Gaussian distribution with zero mean and variance  $\sigma_{h,\ell,k}^2 \frac{P_t}{r_{\ell,k}^\alpha} + \sigma_{v,\ell,k}^{(\psi)2}$ . From (26), we therefore conclude that  $\hat{\mathbf{h}}_{\ell,k}(i)$  has circular Gaussian distribution with zero mean and variance  $\sigma_{h,\ell,k}^2 + \frac{r_{\ell,k}^\alpha}{P_t} \sigma_{v,\ell,k}^{(\psi)2}$ , and  $|\hat{\mathbf{h}}_{\ell,k}(i)|^2$  has exponential distribution with parameter

$$\lambda_{\ell,k} = \frac{1}{\sigma_{h,\ell,k}^2 + \frac{r_{\ell,k}^\alpha}{P_t} \sigma_{v,\ell,k}^{(\psi)2}} \quad (27)$$

From here the probability of successful transmission from node  $\ell$  to node  $k$  will be defined in terms of the estimated channel coefficient as

$$p_{\ell,k} \triangleq \Pr(|\hat{\mathbf{h}}_{\ell,k}(i)|^2 \geq \nu_{\ell,k}) = e^{-\lambda_{\ell,k} \nu_{\ell,k}} \quad (28)$$

Considering the assumed training data and from (25) and (26), the instantaneous channel estimation error will be

$$\tilde{\mathbf{h}}_{\ell,k}(i) = \mathbf{h}_{\ell,k}(i) - \hat{\mathbf{h}}_{\ell,k}(i) = -\sqrt{\frac{r_{\ell,k}^\alpha}{P_t}} \mathbf{v}_{\ell,k}^{(y)}(i) \quad (29)$$

Therefore, the variance of the estimation error is:

$$\sigma_{\tilde{\mathbf{h}}_{\ell,k}}^2 = \mathbb{E}|\tilde{\mathbf{h}}_{\ell,k}(i)|^2 = \frac{r_{\ell,k}^\alpha}{P_t} \sigma_{v,\ell,k}^{(\psi)2} \quad (30)$$

which shows that the power of the channel estimation error,  $\sigma_{\tilde{\mathbf{h}}_{\ell,k}}^2$ , decreases if the node transmit power increases or if the distance between nodes  $\ell$  and  $k$  decreases. To reduce the channel estimation error, the alternative solution is to use more pilot data. It can be shown that if the wireless channel remains invariant over the transmission of  $n$  pilot data, then the estimation error variance will be scaled by a factor of  $1/n$  [41].

**Remark 2.** The time index  $i$ , in Algorithms 1 and 2, refers to the iteration number of adaptation and combination steps and not the time at which the communication between nodes occurs. This implies that from time index  $i - 1$  to  $i$ , a node may transmit several training symbols to its neighbors for channel estimation process and, therefore, the estimated channels used in iteration  $i$  may be obtained using several pilot data. However, to simplify the presentation, we also use index  $i$  to represent the communication time of pilots in (25) since it is assumed that a single pilot datum used for channel estimation.

We can now express (2) in terms of the estimated channels  $\hat{\mathbf{h}}_{\ell,k}(i)$  and the channel estimation error as

$$\boldsymbol{\psi}_{\ell,k,i} = \hat{\mathbf{h}}_{\ell,k}(i) \sqrt{\frac{P_t}{r_{\ell,k}^\alpha}} \boldsymbol{\psi}_{\ell,i} + \sqrt{\frac{P_t}{r_{\ell,k}^\alpha}} \tilde{\mathbf{h}}_{\ell,k}(i) \boldsymbol{\psi}_{\ell,i} + \mathbf{v}_{\ell,k,i}^{(\psi)} \quad (31)$$

The equalization coefficients  $\hat{\mathbf{g}}_{\ell,k}(i)$  are computed using the estimated channels  $\hat{\mathbf{h}}_{\ell,k}(i)$ , according to (16). Using this construction, the equalized received data at node  $k$  become:

$$\hat{\mathbf{g}}_{\ell,k}(i) \boldsymbol{\psi}_{\ell,k,i} = \left(1 + \hat{\mathbf{g}}_{\ell,k}(i) \sqrt{\frac{P_t}{r_{\ell,k}^\alpha}} \tilde{\mathbf{h}}_{\ell,k}(i)\right) \boldsymbol{\psi}_{\ell,i} + \hat{\mathbf{g}}_{\ell,k}(i) \mathbf{v}_{\ell,k,i}^{(\psi)} \quad (32)$$

Substituting the equalized data into (42), we obtain:

$$\mathbf{w}_{k,i} = \sum_{\ell \in \mathcal{N}_{k,i}} \mathbf{a}_{\ell,k}(i) \boldsymbol{\psi}_{\ell,i} + \sum_{\ell \in \mathcal{N}_{k,i}} \mathbf{e}_{\ell,k}(i) \boldsymbol{\psi}_{\ell,i} + \mathbf{v}_{k,i}^{(\psi)} \quad (33)$$

where

$$\mathbf{e}_{\ell,k}(i) = -\mathbf{a}_{\ell,k}(i) \hat{\mathbf{g}}_{\ell,k}(i) \mathbf{v}_{\ell,k,i}^{(\psi)}(i) \quad (34)$$

$$\mathbf{v}_{k,i}^{(\psi)} = \sum_{\ell \in \mathcal{N}_{k,i}} \mathbf{a}_{\ell,k}(i) \hat{\mathbf{g}}_{\ell,k}(i) \mathbf{v}_{\ell,k,i}^{(\psi)} \quad (35)$$

There are several important features in the combination step (33) that need to be highlighted. First, the combination coefficients,  $\mathbf{a}_{\ell,k}(i)$ , used in this step are time varying. These coefficients, in addition to combining the exchanged information, model the link failure phenomenon over the network. Second,  $\{\hat{\mathbf{g}}_{\ell,k}(i)\}$  account for the effects of fading channels. Using these variables and the control SNR mechanism introduced above, we can reduce the effect of link noise. Third, in (33),  $\{\mathbf{e}_{\ell,k}(i)\}$  model the channel estimation errors, which allows us to examine the impact of these errors on the diffusion strategies.

In summary, in a multi-agent wireless network, each node  $k$  will perform the processing tasks listed in Table I in order of precedence to complete cycle  $i$  of the ATC diffusion LMS algorithm.

TABLE I: ATC diffusion implementation

$$\hat{\mathbf{h}}_{\ell,k}(i) = \begin{cases} \sqrt{\frac{r_{\ell,k}^\alpha}{P_t}} \mathbf{y}_{\ell,k}(i) & \text{if } \ell \in \mathcal{N}_k \setminus \{k\} \\ 1 & \text{if } \ell = k \end{cases} \quad (36)$$

$$\mathcal{N}_{k,i} = \left\{ \ell \in \mathcal{N}_k \mid \hat{\mathbf{h}}_{\ell,k}(i) \geq \nu_{\ell,k} \right\} \quad (37)$$

$$\hat{\mathbf{g}}_{\ell,k}(i) = \begin{cases} \frac{\hat{\mathbf{h}}_{\ell,k}^*(i)}{|\hat{\mathbf{h}}_{\ell,k}(i)|^2} \sqrt{\frac{r_{\ell,k}^\alpha}{P_t}} & \text{if } \ell \in \mathcal{N}_{k,i} \setminus \{k\} \\ 1 & \text{if } \ell = k \end{cases} \quad (38)$$

$$\mathcal{I}_{\ell,k}(i) = \begin{cases} 1, & \text{if } \ell \in \mathcal{N}_{k,i} \\ 0, & \text{otherwise} \end{cases} \quad (39)$$

$$\mathbf{a}_{\ell,k}(i) = \begin{cases} \gamma_{\ell,k} \mathcal{I}_{\ell,k}(i), & \text{if } \ell \in \mathcal{N}_{k,i} \setminus \{k\} \\ 1 - \sum_{\ell \in \mathcal{N}_{k,i} \setminus \{k\}} \mathbf{a}_{\ell,k}(i), & \text{if } \ell = k \end{cases} \quad (40)$$

$$\boldsymbol{\psi}_{k,i} = \mathbf{w}_{k,i-1} + \mu_k \mathbf{u}_{k,i}^* [\mathbf{d}_k(i) - \mathbf{u}_{k,i} \mathbf{w}_{k,i-1}] \quad (41)$$

$$\mathbf{w}_{k,i} = \sum_{\ell \in \mathcal{N}_{k,i}} \mathbf{a}_{\ell,k}(i) \hat{\mathbf{g}}_{\ell,k}(i) \boldsymbol{\psi}_{\ell,k,i} \quad (42)$$

## 4 PERFORMANCE ANALYSIS

In this section, we derive conditions under which the equalized diffusion strategies are stable in the mean and mean square sense. We also derive expressions to characterize the mean-square-deviation (MSD) and excess mean-square-error (EMSE) performance levels of the algorithms during the transient phase and in steady-state. We focus on the ATC variant (41)–(42). The same conclusions hold for (17)–(18) with minor adjustments.

To derive a recursion for the mean error-vector of the network, we begin with defining the local error vectors:

$$\tilde{\mathbf{w}}_{k,i} \triangleq \mathbf{w}^o - \mathbf{w}_{k,i} \quad (43)$$

$$\tilde{\boldsymbol{\psi}}_{k,i} \triangleq \mathbf{w}^o - \boldsymbol{\psi}_{k,i} \quad (44)$$

We subtract  $\mathbf{w}^o$  from both sides of (41) and (33) to obtain:

$$\tilde{\boldsymbol{\psi}}_{k,i} = (I - \mu_k \mathbf{u}_{k,i}^* \mathbf{u}_{k,i}) \tilde{\mathbf{w}}_{k,i-1} - \mu_k \mathbf{u}_{k,i}^* \mathbf{v}_k(i) \quad (45)$$

$$\begin{aligned} \tilde{\mathbf{w}}_{k,i} = & \sum_{\ell \in \mathcal{N}_{k,i}} \mathbf{a}_{\ell,k}(i) \tilde{\boldsymbol{\psi}}_{\ell,i} + \sum_{\ell \in \mathcal{N}_{k,i}} \mathbf{e}_{\ell,k}(i) \tilde{\boldsymbol{\psi}}_{\ell,i} \\ & + \sum_{\ell \in \mathcal{N}_{k,i}} \mathbf{e}_{\ell,k}(i) \mathbf{w}^o - \mathbf{v}_{k,i}^{(\psi)} \end{aligned} \quad (46)$$

We collect the  $\{\mathbf{a}_{\ell,k}(i)\}$  into a left-stochastic matrix  $\mathbf{A}_i$  and the  $\{\mathbf{e}_{\ell,k}(i)\}$  into an error matrix  $\mathbf{E}_i$ . We also define the extended versions of these matrices using Kronecker products as  $\mathcal{A}_i \triangleq \mathbf{A}_i \otimes I_M$  and  $\mathcal{E}_i \triangleq \mathbf{E}_i \otimes I_M$ . We further introduce the network error vectors:

$$\tilde{\boldsymbol{\psi}}_i \triangleq \text{col}\{\tilde{\boldsymbol{\psi}}_{1,i}, \tilde{\boldsymbol{\psi}}_{2,i}, \dots, \tilde{\boldsymbol{\psi}}_{N,i}\} \quad (47)$$

$$\tilde{\mathbf{w}}_i \triangleq \text{col}\{\tilde{\mathbf{w}}_{1,i}, \tilde{\mathbf{w}}_{2,i}, \dots, \tilde{\mathbf{w}}_{N,i}\} \quad (48)$$

and the variables:

$$\mathcal{R}_i \triangleq \text{diag}\{\mathbf{u}_{1,i}^* \mathbf{u}_{1,i}, \dots, \mathbf{u}_{N,i}^* \mathbf{u}_{N,i}\} \quad (49)$$

$$\mathcal{M} \triangleq \text{diag}\{\mu_1 I_M, \dots, \mu_N I_M\} \quad (50)$$

$$\mathbf{p}_i \triangleq \text{col}\{\mathbf{u}_{1,i}^* \mathbf{v}_1(i), \dots, \mathbf{u}_{N,i}^* \mathbf{v}_N(i)\} \quad (51)$$

$$\mathbf{v}_i^{(\psi)} \triangleq \text{col}\{\mathbf{v}_{1,i}^{(\psi)}, \dots, \mathbf{v}_{N,i}^{(\psi)}\} \quad (52)$$

$$\boldsymbol{\omega}^o \triangleq \mathbf{1}_N \otimes \mathbf{w}^o \quad (53)$$

where  $\mathbf{1}_N$  is a column vector with length  $N$  and unit entries. We can now use (45) and (46) to verify that the following recursion holds for the network error vector:

$$\tilde{\mathbf{w}}_i = \mathcal{B}_i \tilde{\mathbf{w}}_{i-1} - (\mathcal{A}_i + \mathcal{E}_i)^T \mathcal{M} \mathbf{p}_i + \mathcal{E}_i^T \boldsymbol{\omega}^o - \mathbf{v}_i^{(\psi)} \quad (54)$$

where

$$\mathcal{B}_i = (\mathcal{A}_i + \mathcal{E}_i)^T (I - \mathcal{M} \mathcal{R}_i) \quad (55)$$

### 4.1 Mean Convergence

Taking the expectation of (54) under Assumptions 1 and 2, we arrive at

$$\mathbb{E}[\tilde{\mathbf{w}}_i] = \mathcal{B} \mathbb{E}[\tilde{\mathbf{w}}_{i-1}] + \mathcal{E}^T \boldsymbol{\omega}^o \quad (56)$$

where

$$\mathcal{B} \triangleq \mathbb{E}[\mathcal{B}_i] = (\mathcal{A} + \mathcal{E})^T (I - \mathcal{M} \mathcal{R}) \quad (57)$$

$$\mathcal{A} \triangleq \mathbb{E}[\mathcal{A}_i] = \mathcal{A} \otimes I_M \quad (58)$$

$$\mathcal{E} \triangleq \mathbb{E}[\mathcal{E}_i] = \mathcal{E} \otimes I_M \quad (59)$$

$$\mathcal{R} \triangleq \mathbb{E}[\mathcal{R}_i] = \text{diag}\{R_{u,1}, \dots, R_{u,N}\} \quad (60)$$

To obtain (56), we used the fact that  $\mathbf{v}_k(i)$  is independent of  $\mathbf{u}_{k,i}$  and  $\mathbb{E}[\mathbf{v}_k(i)] = 0$ . Moreover, we have  $\mathbb{E}[\mathbf{v}_i^{(\psi)}] = 0$  because  $\hat{\mathbf{g}}_{\ell,k}(i)$  is independent of  $\mathbf{v}_{\ell,k,i}^{(\psi)}$  and  $\mathbb{E}[\mathbf{v}_{\ell,k,i}^{(\psi)}] = 0$ . Considering the time-varying left-stochastic matrix  $\mathbf{A}_i$ , we can use (22) to find the entries of  $\mathcal{A} = \mathbb{E}[\mathcal{A}_i]$ , i.e.,

$$a_{\ell,k} = \begin{cases} \gamma_{\ell,k} p_{\ell,k}, & \text{if } \ell \in \mathcal{N}_k \setminus \{k\} \\ 1 - \sum_{\ell \in \mathcal{N}_k \setminus \{k\}} \gamma_{\ell,k} p_{\ell,k}, & \text{if } \ell = k \end{cases} \quad (61)$$

Observe that  $A^T \mathbf{1} = \mathbf{1}$ . The  $(\ell, k)$ -th entry of matrix  $E$  is zero on the diagonal and, for  $\ell \neq k$ , is given by:

$$\begin{aligned}
e_{\ell,k} &= -\mathbb{E}[\mathbf{a}_{\ell,k}(i) \hat{\mathbf{g}}_{\ell,k}(i) \mathbf{v}_{\ell,k}^{(y)}(i)] \\
&\stackrel{(i)}{=} -\gamma_{\ell,k} \mathbb{E}[\mathcal{I}_{\ell,k}(i) \hat{\mathbf{g}}_{\ell,k}(i) \mathbf{v}_{\ell,k}^{(y)}(i)] \\
&\stackrel{(ii)}{=} -\gamma_{\ell,k} \mathbb{E}[\hat{\mathbf{g}}_{\ell,k}(i) \mathbf{v}_{\ell,k}^{(y)}(i) | \hat{\mathbf{h}}_{\ell,k}(i)|^2 \geq \nu_{\ell,k}] \\
&\stackrel{(iii)}{=} -\gamma_{\ell,k} \mathbb{E} \left[ \frac{\left( \sqrt{\frac{r^\alpha}{P_t}} \mathbf{h}_{\ell,k}^*(i) \mathbf{v}_{\ell,k}^{(y)}(i) + \frac{r^\alpha}{P_t} |\mathbf{v}_{\ell,k}^{(y)}(i)|^2 \right)}{|\mathbf{h}_{\ell,k}(i) + \sqrt{\frac{r^\alpha}{P_t}} \mathbf{v}_{\ell,k}^{(y)}(i)|^2} \right. \\
&\quad \left. \left| \left( |\mathbf{h}_{\ell,k}(i) + \sqrt{\frac{r^\alpha}{P_t}} \mathbf{v}_{\ell,k}^{(y)}(i)|^2 \geq \nu_{\ell,k} \right) \right] \quad (62)
\end{aligned}$$

The equality in step (ii) follows from the fact that  $\hat{\mathbf{g}}_{\ell,k}(i)$  is defined for  $\ell \in \mathcal{N}_k \setminus \{k\}$  when  $|\hat{\mathbf{h}}_{\ell,k}(i)|^2 \geq \nu_{\ell,k}$ , for which  $\mathcal{I}_{\ell,k}(i) = 1$ . We obtain (iii) by expressing  $\hat{\mathbf{g}}_{\ell,k}(i)$  in terms of  $\mathbf{h}_{\ell,k}(i)$  and  $\mathbf{v}_{\ell,k}^{(y)}(i)$  according to (25), (36) and (38). Expression (62) indicates that  $e_{\ell,k}$  is bounded.

**Remark 3.** From the right hand side of (62), it can be verified that the value of the expectation is independent of time since the estimation error,  $\mathbf{v}_{\ell,k}^{(y)}(i)$ , and the channel coefficients,  $\mathbf{h}_{\ell,k}(i)$ , are assumed to be i.i.d. over time with fixed probability density functions.

According to (56), when  $\mathcal{B}$  is stable, then the network mean error vector converges to

$$\mathbf{b} \triangleq \lim_{i \rightarrow \infty} \mathbb{E}[\tilde{\mathbf{w}}_i] = (\mathbf{I} - \mathcal{B})^{-1} \mathcal{E}^T \omega^o \quad (63)$$

If  $\hat{\mathbf{h}}_{\ell,k}(i) = \mathbf{h}_{\ell,k}(i)$  then  $\mathcal{E} = 0$  and  $\lim_{i \rightarrow \infty} \mathbb{E}[\tilde{\mathbf{w}}_i] = 0$ , i.e., the algorithm will be asymptotically unbiased.

Let us now find conditions under which  $\mathcal{B}$  is stable, i.e., conditions under which the spectral radius of  $\mathcal{B}$ , denoted by  $\rho(\mathcal{B})$ , is strictly less than one. We use the properties of the block maximum norm  $\|\cdot\|_{b,\infty}$  from [3], [42] to establish the following relations:

$$\begin{aligned}
\rho(\mathcal{B}) &\leq \|\mathcal{B}\|_{b,\infty} \\
&\leq \|(\mathcal{A} + \mathcal{E})^T\|_{b,\infty} \|(\mathbf{I} - \mathcal{M}\mathcal{R})\|_{b,\infty} \\
&\leq \left( \|\mathcal{A}^T\|_{b,\infty} + \|\mathcal{E}^T\|_{b,\infty} \right) \|(\mathbf{I} - \mathcal{M}\mathcal{R})\|_{b,\infty} \\
&= (1 + \|\mathcal{E}^T\|_{b,\infty}) \|(\mathbf{I} - \mathcal{M}\mathcal{R})\|_{b,\infty} \quad (64)
\end{aligned}$$

where in the last equality we used the fact that  $\|\mathcal{A}^T\|_{b,\infty} = 1$  since  $A$  is left-stochastic. According to (64),  $\rho(\mathcal{B})$  is bounded by one if

$$\|(\mathbf{I} - \mathcal{M}\mathcal{R})\|_{b,\infty} < \frac{1}{1 + \|\mathcal{E}\|_{b,\infty}} \quad (65)$$

Since  $\mathbf{I} - \mathcal{M}\mathcal{R}$  is block diagonal and Hermitian, we have  $\|(\mathbf{I} - \mathcal{M}\mathcal{R})\|_{b,\infty} = \rho(\mathbf{I} - \mathcal{M}\mathcal{R})$  [3]. The spectral radius of  $\mathbf{I} - \mathcal{M}\mathcal{R}$  will be less than  $1/(1 + \|\mathcal{E}\|_{b,\infty})$  if the absolute maximum eigenvalue of each of its blocks is strictly less than  $1/(1 + \|\mathcal{E}\|_{b,\infty})$ . This condition is satisfied if at each node  $k$  the step-size  $\mu_k$  is chosen as:

$$\frac{1 - \frac{1}{1 + \|\mathcal{E}\|_{b,\infty}}}{\lambda_{\max}(R_{u,k})} < \mu_k < \frac{1 + \frac{1}{1 + \|\mathcal{E}\|_{b,\infty}}}{\lambda_{\max}(R_{u,k})} \quad (66)$$

where  $\lambda_{\max}(\cdot)$  denotes the maximum eigenvalue of its matrix argument. This relation reveals that the mean-stability range of the algorithm, in terms of the step size parameters  $\{\mu_k\}$ , reduces as the channel estimation error over the network increases. When the channel estimation error approaches zero<sup>1</sup>, that is when  $\|\mathcal{E}\|_{b,\infty} \rightarrow 0$ , the stability condition reduces to  $0 < \mu_k < \frac{2}{\lambda_{\max}(R_{u,k})}$ , which is the mean stability range of diffusion LMS over ideal communication links [3]. A similar analysis can be carried out for the CTA diffusion strategy.

**Theorem 1.** Consider the diffusion strategies (41)–(42) with the space-time data (1) and (2) satisfying Assumptions 1 and 2, respectively, and where the channel coefficients are estimated using (26) with training symbols  $s_k(i) = 1$ . Then the algorithms will be stable in the mean and the mean error vector will converge to (63) if the step-sizes are chosen according to (66).

## 4.2 Steady-State Mean-Square Performance

To study the mean-square performance of the algorithm, we need to determine the network variance relation [1], [26], [43]. The latter can be obtained by equating the weighted squared norms of both sides of (54), and taking expectations under Assumptions 1 and 2:

$$\begin{aligned}
\mathbb{E}\|\tilde{\mathbf{w}}_i\|_\Sigma^2 &= \mathbb{E}\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'_i}^2 + \mathbb{E}[\omega^{o*} \mathcal{E}_i^{*T} \Sigma \mathcal{E}_i^T \omega^o] \\
&\quad + \mathbb{E}[\mathbf{p}_i^* \mathcal{M}^T (\mathcal{A}_i + \mathcal{E}_i^{*T}) \Sigma (\mathcal{A}_i + \mathcal{E}_i)^T \mathcal{M} \mathbf{p}_i] \\
&\quad + 2\text{Re}\left\{ \mathbb{E}[\omega^{o*} \mathcal{E}_i^{*T} \Sigma \mathcal{B}_i \tilde{\mathbf{w}}_{i-1}] \right\} + \mathbb{E}[\mathbf{v}_i^{(\psi)*} \Sigma \mathbf{v}_i^{(\psi)}] \quad (67)
\end{aligned}$$

where for a vector  $x$  and a weighting matrix  $\Sigma \geq 0$  with compatible dimensions  $\|x\|_\Sigma^2 = x^* \Sigma x$ , and

$$\Sigma'_i = \mathcal{B}_i^* \Sigma \mathcal{B}_i \quad (68)$$

Under the independence assumption between  $\tilde{\mathbf{w}}_{i-1}$  and  $\mathcal{R}_i$ , it holds that

$$\mathbb{E}[\|\tilde{\mathbf{w}}_{i-1}\|_{\Sigma'_i}^2] = \mathbb{E}\|\tilde{\mathbf{w}}_{i-1}\|_{\mathbb{E}[\Sigma'_i]}^2 \quad (69)$$

Using this equality in (67), we arrive at:

$$\begin{aligned}
\mathbb{E}\|\tilde{\mathbf{w}}_i\|_\Sigma^2 &= \mathbb{E}\|\tilde{\mathbf{w}}_{i-1}\|_\Sigma^2 + \text{Tr}(\mathbb{E}[\mathcal{E}_i^T \omega^{o*} \omega^o \mathcal{E}_i^{*T} \Sigma]) \\
&\quad + \text{Tr}(\mathbb{E}[(\mathcal{A}_i + \mathcal{E}_i)^T \mathcal{M} \mathbf{p}_i \mathbf{p}_i^* \mathcal{M} (\mathcal{A}_i + \mathcal{E}_i^{*T}) \Sigma]) \\
&\quad + 2\text{Re}\left\{ \text{Tr}(\mathbb{E}[\mathcal{B}_i \tilde{\mathbf{w}}_{i-1} \omega^{o*} \mathcal{E}_i^{*T} \Sigma]) \right\} + \text{Tr}(\mathbb{E}[\mathbf{v}_i^{(\psi)} \mathbf{v}_i^{(\psi)*} \Sigma]) \quad (70)
\end{aligned}$$

where  $\Sigma' = \mathbb{E}[\Sigma'_i]$ . To compute (70), we introduce:

$$\mathcal{P} \triangleq \mathbb{E}[\mathbf{p}_i \mathbf{p}_i^*] = \text{diag}\{\sigma_{v,1}^2 R_{u,1}, \dots, \sigma_{v,N}^2 R_{u,N}\} \quad (71)$$

$$\mathcal{R}_v \triangleq \text{diag}\{R_{v,1}, \dots, R_{v,N}\} \quad (72)$$

$$R_{v,k} \triangleq \mathbb{E}[\mathbf{v}_{k,i}^{(\psi)} \mathbf{v}_{k,i}^{(\psi)*}] = \sum_{\ell \in \mathcal{N}_k \setminus \{k\}} \mathbb{E}[\mathbf{a}_{\ell,k}^2(i) |\hat{\mathbf{g}}_{\ell,k}(i)|^2] R_{v,\ell k}^{(\psi)} \quad (73)$$

We show in Appendix A how to compute the expectation term multiplying  $R_{v,\ell k}^{(\psi)}$  in (73). Alternatively, this

1. The channel estimation error can be reduced by transmitting more pilot symbols or increasing the SNR during pilot transmission.

term can be evaluated numerically by averaging over repeated independent experiments.

To proceed, we assume that  $\Sigma$  is partitioned into block entries of size  $M \times M$  and let  $\sigma = \text{bvec}(\Sigma)$  denote the vector that is obtained from the block vectorization of  $\Sigma$ . We shall write  $\|\tilde{\mathbf{w}}_i\|_\Sigma^2$  and  $\|\tilde{\mathbf{w}}_i\|_\sigma^2$  interchangeably to denote the same weighted square norm [1]. Using properties of  $\text{bvec}$  and block Kronecker products [44], the variance relation in (70) leads in steady-state to:

$$\lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i\|_\sigma^2 = \lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_{i-1}\|_{\mathcal{F}\sigma}^2 + \gamma^T \sigma \quad (74)$$

where  $\mathcal{F} = \mathbb{E}[\mathcal{B}_i^T \otimes_b \mathcal{B}_i^*]$ , and

$$\begin{aligned} \gamma &= \lim_{i \rightarrow \infty} \left\{ \mathbb{E}[\mathcal{E}_i^T \otimes_b \mathcal{E}_i^*] \text{bvec}((\omega^o \omega^{o*})^T) \right. \\ &\quad + \mathbb{E}[(\mathcal{A}_i + \mathcal{E}_i)^T \otimes_b (\mathcal{A}_i + \mathcal{E}_i^{*T})] \text{bvec}(\mathcal{M}\mathcal{P}^T\mathcal{M}) \\ &\quad \left. + 2\text{Re}\{\mathbb{E}[\mathcal{B}_i \otimes_b \mathcal{E}_i^*] \text{bvec}((b\omega^{o*})^T)\} + \text{bvec}(R_v^T) \right\} \quad (75) \end{aligned}$$

Considering (55), matrix  $\mathcal{F}$  can be written as:

$$\begin{aligned} \mathcal{F} &= \mathbb{E} \left\{ [(I - \mathcal{M}\mathcal{R}_i)^T (\mathcal{A}_i + \mathcal{E}_i)] \right. \\ &\quad \left. \otimes_b [(\mathcal{A}_i + \mathcal{E}_i^{*T})(I - \mathcal{M}\mathcal{R}_i)] \right\} \\ &= \mathbb{E} \left\{ [(I - \mathcal{M}\mathcal{R}_i)^T \otimes_b (I - \mathcal{M}\mathcal{R}_i)] \right. \\ &\quad \left. \times [(\mathcal{A}_i + \mathcal{E}_i) \otimes_b (\mathcal{A}_i + \mathcal{E}_i^{*T})] \right\} \quad (76) \end{aligned}$$

Since the entries of matrix  $\mathcal{R}_i$ , which are defined in terms of the regression data  $\mathbf{u}_{k,i}$ , are independent of the entries of matrices  $\mathcal{A}_i$  and  $\mathcal{E}_i$ , i.e.,  $\mathbf{a}_{\ell,k}(i)$  and  $\mathbf{e}_{\ell,k}(i)$ , matrix  $\mathcal{F}$  in (76) can be written more compactly as:

$$\mathcal{F} = \bar{\mathcal{F}}\mathcal{D} \quad (77)$$

where

$$\bar{\mathcal{F}} \triangleq \mathbb{E} [(I - \mathcal{M}\mathcal{R}_i)^T \otimes_b (I - \mathcal{M}\mathcal{R}_i)] \quad (78)$$

$$\mathcal{D} \triangleq \mathbb{E}[\mathcal{D}_i] = \mathbb{E} [(\mathcal{A}_i + \mathcal{E}_i) \otimes_b (\mathcal{A}_i + \mathcal{E}_i^{*T})] \quad (79)$$

We can find an expression for  $\bar{\mathcal{F}}$  if we assume that the regression data  $\mathbf{u}_{k,i}$  are circular Gaussian—see equation (80) and Appendix B, where  $\mathbf{e}_k$  is a unit basis vector in  $\mathbb{R}^N$  with entry one at position  $k$ ,  $r_k = \text{vec}(R_{u,k})$ ,  $\beta = 2$  for real-valued data and  $\beta = 1$  for complex-valued data. A simplified expression can be found to compute  $\bar{\mathcal{F}}$  without using the Gaussian assumption on the regression data provided that the following condition holds.

**Assumption 3.** *The channel estimation errors over the network are small enough such that the adaptation step-sizes in (66) can be chosen sufficiently small.*

In cases where the distribution of the regression data is unknown, under Assumption 3, the contributing terms

depending on  $\mu_k^2$  can be neglected and as a result  $\bar{\mathcal{F}}$  can be approximated by

$$\bar{\mathcal{F}} \approx [(I - \mathcal{M}\mathcal{R})^T \otimes_b (I - \mathcal{M}\mathcal{R})] \quad (81)$$

In Appendix C, we show how to obtain the matrix  $\mathcal{D}$  in (79) needed for computing  $\mathcal{F}$  in (77). To evaluate  $\gamma$ , we use the following relations, which are also established in Appendix C:

$$\mathbb{E}[\mathcal{E}_i^T \otimes_b \mathcal{E}_i^*] = \mathbb{E}[(\mathbf{E}_i^T \otimes \mathbf{E}_i^*)] \otimes I_{M^2} \quad (82)$$

$$\begin{aligned} \mathbb{E}[(\mathcal{A}_i + \mathcal{E}_i)^T \otimes_b (\mathcal{A}_i + \mathcal{E}_i^*)] \\ = \left( \mathbb{E}[(\mathbf{A}_i \otimes \mathbf{A}_i)^T] + \mathbb{E}[(\mathbf{A}_i^T \otimes \mathbf{E}_i^*)] \right. \\ \left. + \mathbb{E}[(\mathbf{E}_i \otimes \mathbf{A}_i)^T] + \mathbb{E}[(\mathbf{E}_i^T \otimes \mathbf{E}_i^*)] \right) \otimes I_{M^2} \quad (83) \end{aligned}$$

$$\begin{aligned} \mathbb{E}[\mathcal{B}_i \otimes_b \mathcal{E}_i^*] &= \left\{ \left( \mathbb{E}[(\mathbf{A}_i \otimes \mathbf{E}_i)^*] + \mathbb{E}[(\mathbf{E}_i^T \otimes \mathbf{E}_i^*)] \right) \otimes I_{M^2} \right\} \\ &\quad \times \left\{ (I_{MN} - \mathcal{M}\mathcal{R}) \otimes_b I_{MN} \right\} \quad (84) \end{aligned}$$

To obtain mean-square error (MSE) steady state expressions for the network, we let  $i$  go to infinity and use expression (74) to write:

$$\lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{(I-\mathcal{F})\sigma}^2 = \gamma^T \sigma \quad (85)$$

Since we are free to choose  $\Sigma$  and hence  $\sigma$ , we choose  $(I - \mathcal{F})\sigma = \text{bvec}(\Omega)$ , where  $\Omega$  is another arbitrary positive semidefinite matrix. Doing so, we arrive at:

$$\lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i\|_\Omega^2 = \gamma^T (I - \mathcal{F})^{-1} \text{bvec}(\Omega) \quad (86)$$

Recall from (48) that each sub-vector of  $\tilde{\mathbf{w}}_i$  corresponds to the estimation error at a particular node, for instance,  $\tilde{\mathbf{w}}_{k,i}$  is the estimation error at node  $k$ . Therefore, using (86), the MSD at node  $k$ , denoted by  $\eta_k$ , can be computed by choosing  $\Omega = \{\text{diag}(\mathbf{e}_k) \otimes I\}$ , i.e.:

$$\begin{aligned} \eta_k &= \lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|^2 = \lim_{i \rightarrow \infty} \mathbb{E} \|\tilde{\mathbf{w}}_i\|_{\{\text{diag}(\mathbf{e}_k) \otimes I\}}^2 \\ &= \gamma^T (I - \mathcal{F})^{-1} \text{bvec}(\text{diag}(\mathbf{e}_k) \otimes I_M) \quad (87) \end{aligned}$$

The network MSD, denoted by  $\eta$ , is then defined as:

$$\eta = \lim_{i \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \mathbb{E} \|\tilde{\mathbf{w}}_{k,i}\|^2 \quad (88)$$

which it can be also computed from (86) by using  $\Omega = \frac{1}{N} I_{MN}$ . This leads to:

$$\eta = \lim_{i \rightarrow \infty} \frac{1}{N} \mathbb{E} \|\tilde{\mathbf{w}}_i\|^2 = \frac{1}{N} \gamma^T (I - \mathcal{F})^{-1} \text{bvec}(I_{MN}) \quad (89)$$

In (87) and (89), we assume that  $(I - \mathcal{F})$  is invertible. In what follows, we find conditions under which this assumption is satisfied. Using the properties of the Kronecker product and the sub-multiplicative property of norms, we can write:

$$\rho(\mathcal{F}) \leq \|\bar{\mathcal{F}}\mathcal{D}\|_{b,\infty} \leq \|\bar{\mathcal{F}}\|_{b,\infty} \|\mathcal{D}\|_{b,\infty} \quad (90)$$

---


$$\bar{\mathcal{F}} = (I - \mathcal{M}\mathcal{R})^T \otimes_b (I - \mathcal{M}\mathcal{R}) + \left\{ \sum_{k=1}^N \left[ \text{diag}((\text{vec}(\text{diag}(\mathbf{e}_k))) \right] \otimes [(\beta - 1)(R_{k,u}^T \otimes R_{k,u}) + r_k r_k^*] \right\} (\mathcal{M} \otimes_b \mathcal{M}) \quad (80)$$



We next show that  $\bar{\mathcal{F}}$  from (81) is a block diagonal Hermitian matrix with block size  $NM^2 \times NM^2$ . To this end, we note that  $I - \mathcal{MR}$  is a block diagonal matrix with block size  $M \times M$  and then use (81) to obtain:

$$\bar{\mathcal{F}} = \text{diag}\left\{(I - \mu_1 R_{u,1})^T \otimes (I - \mathcal{MR}), \dots, (I - \mu_N R_{u,N})^T \otimes (I - \mathcal{MR})\right\} \quad (91)$$

Moreover,  $\bar{\mathcal{F}}$  is Hermitian because considering  $\mathcal{R} = \mathcal{R}^*$ ,  $\mathcal{M} = \mathcal{M}^T$ ,  $\mathcal{R}\mathcal{M} = \mathcal{MR}$ , we will have

$$\begin{aligned} \bar{\mathcal{F}}^* &= ((I - \mathcal{MR})^T)^* \otimes_b (I - \mathcal{MR})^* \\ &= (I - \mathcal{MR})^T \otimes_b (I - \mathcal{MR}) = \bar{\mathcal{F}} \end{aligned} \quad (92)$$

Now we can use the following lemma to bound the spectral radius of matrix  $\mathcal{F}$  in (90).

**Lemma 1.** Consider an  $N \times N$  block diagonal Hermitian matrix  $Y = \text{diag}\{Y_1, Y_2, \dots, Y_N\}$ , where each block  $Y_k$  is of size  $M \times M$  and Hermitian. Then it holds that [3]:

$$\|Y\|_{b,\infty} = \max_{1 \leq k \leq N} \rho(Y_k) = \rho(Y) \quad (93)$$

According to this lemma, since  $\bar{\mathcal{F}}$  is block diagonal Hermitian, we can substitute its block maximum norm on the right hand side of relation (90) with its spectral radius and obtain:

$$\begin{aligned} \rho(\mathcal{F}) &\leq \rho\left((I - \mathcal{MR})^T \otimes_b (I - \mathcal{MR})\right) \|\mathcal{D}\|_{b,\infty} \\ &= \rho^2(I - \mathcal{MR}) \|\mathcal{D}\|_{b,\infty} \end{aligned} \quad (94)$$

We then deduce that  $\rho(\mathcal{F}) < 1$  if:

$$0 < \rho(I - \mathcal{MR}) < \frac{1}{\sqrt{\|\mathcal{D}\|_{b,\infty}}} \quad (95)$$

Since  $I - \mathcal{MR}$  is a block-diagonal matrix, this condition will be satisfied for small step-sizes that also satisfy:

$$\frac{1 - \frac{1}{\sqrt{\|\mathcal{D}\|_{b,\infty}}}}{\lambda_{\max}(R_{u,k})} < \mu_k < \frac{1 + \frac{1}{\sqrt{\|\mathcal{D}\|_{b,\infty}}}}{\lambda_{\max}(R_{u,k})} \quad (96)$$

If the channel estimation error is small, then  $\|\mathcal{E}\|_{b,\infty} \approx 0$  and  $\mathcal{D} \approx \mathcal{A} \otimes_b \mathcal{A}$ . Subsequently,  $\|\mathcal{D}\|_{b,\infty} \approx 1$  and this mean-square stability condition reduces to  $0 < \mu_k < \frac{2}{\lambda_{\max}(R_{u,k})}$  which is the mean-square stability range of diffusion LMS over ideal communication links [3].

### 4.3 Mean-Square Transient Behavior

In this part, we derive expressions to characterize the mean-square convergence behavior of the diffusion algorithms over wireless networks with fading channels and noisy communication links. To derive these expressions, it is assumed that each node knows the CSI of its neighbors, and  $\mathbf{E}_i = 0$  for all  $i$ . We then use (67) and consider  $\mathbf{w}_{k,-1} = 0$ ,  $\forall k \in \{1, \dots, N\}$  to arrive at:

$$\mathbb{E}\|\tilde{\mathbf{w}}_i\|_\sigma^2 = \|\mathbf{w}^o\|_{\mathcal{F}^{i+1}\sigma}^2 + \bar{\gamma}^T \sum_{j=0}^i \mathcal{F}^j \sigma \quad (97)$$

where

$$\bar{\gamma} \triangleq \mathbb{E}[\mathcal{A}_i^T \otimes_b \mathcal{A}_i] \text{bvec}(\mathcal{MP}^T \mathcal{M}) + \text{bvec}(\bar{R}_v^T) \quad (98)$$

$$\bar{R}_v \triangleq \text{diag}\{\bar{R}_{v,1}, \dots, \bar{R}_{v,N}\} \quad (99)$$

$$\bar{R}_{v,k} \triangleq \mathbb{E}[\mathbf{v}_{k,i}^{(\psi)} \mathbf{v}_{k,i}^{(\psi)*}] = \sum_{\ell \in \mathcal{N}_k \setminus \{k\}} \mathbb{E}[\mathbf{a}_{\ell,k}^2(i) |\mathbf{g}_{\ell,k}(i)|^2] R_{v,\ell,k}^{(\psi)} \quad (100)$$

Under this condition, and since  $\mathbf{E}_i = 0$ ,  $\mathcal{F}$  can be expressed as:

$$\mathcal{F} \approx \bar{\mathcal{F}} \mathbb{E}[\mathcal{A}_i^T \otimes_b \mathcal{A}_i] \quad (101)$$

Writing (97) for  $i-1$  and computing  $\mathbb{E}\|\tilde{\mathbf{w}}_i\|_\sigma^2 - \mathbb{E}\|\tilde{\mathbf{w}}_{i-1}\|_\sigma^2$  leads to:

$$\mathbb{E}\|\tilde{\mathbf{w}}_i\|_\sigma^2 = \mathbb{E}\|\tilde{\mathbf{w}}_{i-1}\|_\sigma^2 + \|\mathbf{w}^o\|_{\mathcal{F}^i(I-\mathcal{F})\sigma}^2 + \bar{\gamma}^T \mathcal{F}^i \sigma \quad (102)$$

By replacing  $\sigma$  with  $\sigma_{\text{msd}_k} = \text{diag}(e_k) \otimes I_M$  and  $\sigma_{\text{emse}_k} = \text{diag}(e_k) \otimes R_{u,k}$ , we arrive at two recursions for the evolution of the MSD and EMSE over time:

$$\eta_k(i) = \eta_k(i-1) - \|\mathbf{w}^o\|_{\mathcal{F}^i(I-\mathcal{F})\sigma_{\text{msd}_k}}^2 + \bar{\gamma}^T \mathcal{F}^i \sigma_{\text{msd}_k} \quad (103)$$

$$\zeta_k(i) = \zeta_k(i-1) - \|\mathbf{w}^o\|_{\mathcal{F}^i(I-\mathcal{F})\sigma_{\text{emse}_k}}^2 + \bar{\gamma}^T \mathcal{F}^i \sigma_{\text{emse}_k} \quad (104)$$

We can find the learning curves of the network MSD and EMSE either by averaging the nodes learning curves (103) and (104), or by, respectively, substituting the following two values for  $\sigma$  in recursion (102):

$$\sigma_{\text{msd}} = \frac{1}{N} \text{bvec}(I_{MN}) \quad (105)$$

$$\sigma_{\text{emse}} = \frac{1}{N} \text{bvec}(\text{diag}\{R_{u,1}, \dots, R_{u,N}\}) \quad (106)$$

## 5 NUMERICAL RESULTS

In this section, we present computer experiments to illustrate the performance of the ATC diffusion strategy (41)–(42) in the estimation of the unknown parameter vector  $\mathbf{w}^o = 2[1+j1, -1+j1]^T$  over time-varying wireless channels. We consider a network with  $N = 10$  nodes, which are randomly spread over a unit square area  $(x, y) \in [0, 1] \times [0, 1]$ , as shown Fig. 2. We choose the transmit power of  $P_t = 1$ , nominal transmission range of  $r_o = 0.4$  and the path-loss exponents  $\alpha = 3.2$ . For each node  $k \in \{1, 2, \dots, N\}$ , we set  $\mu_k = 0.01$  and  $\mathbf{w}_{k,-1} = 0$ . We adopt zero-mean Gaussian random distributions to generate  $\mathbf{v}_k(i)$ ,  $\mathbf{v}_{\ell k,i}^{(\psi)}$  and  $\mathbf{u}_{k,i}$ . The distribution of the communication noise power over the spatial domain is illustrated in Fig. 3. The regression data  $\mathbf{u}_{k,i}$  have covariance matrices of the form  $R_{u,k} = \sigma_{u,k}^2 I_M$ . The trace of the regression data,  $\text{Tr}(R_{u,k})$ , and the variances of measurement noise,  $\sigma_{v,k}^2$ , are illustrated in Fig. 4.

The exchanged data between nodes experience distortion characterized by (2). At time  $i$ , the link between nodes  $\ell$  and  $k$  fails with probability  $1 - p_{\ell,k}$ . We obtain  $\gamma_{\ell,k}$  using the relative-degree combination rule [2], [3], i.e.,

$$\gamma_{\ell,k} = \begin{cases} \frac{|\mathcal{N}_\ell|}{\sum_{m \in \mathcal{N}_k} |\mathcal{N}_m|}, & \text{if } \ell \in \mathcal{N}_k \\ 0, & \text{otherwise} \end{cases} \quad (107)$$

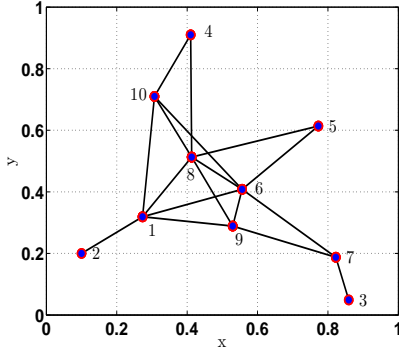


Fig. 2: This graph shows the topology of the wireless network at the start-up time  $i = 0$ , where two nodes are connected if their distance is less than their transmission range,  $r_o = 0.4$ .

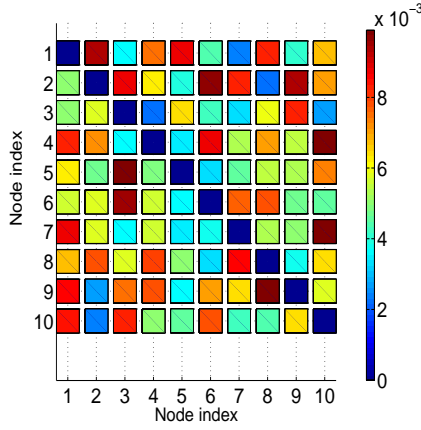


Fig. 3: Power of communication noise  $v_{\ell k, i}^{(\psi)}$  over the network.

and update  $A_i$  it at each time  $i$  according to the introduced combination rule (22).

Figures 5 and 6 show the network MSD in transient and steady-state regimes, where the simulation curves are obtained from the average of 500 independent runs. In these figures, we compare the performance of the proposed ATC diffusion algorithm over wireless channels for different CSI cases at the receiving nodes. In particular, we examine the performance of the algorithm with perfect CSI, where each node  $k$  knows the CSI of all its neighbors. We also consider scenarios where nodes do

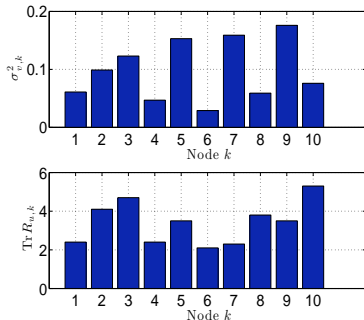


Fig. 4: Network energy profile.

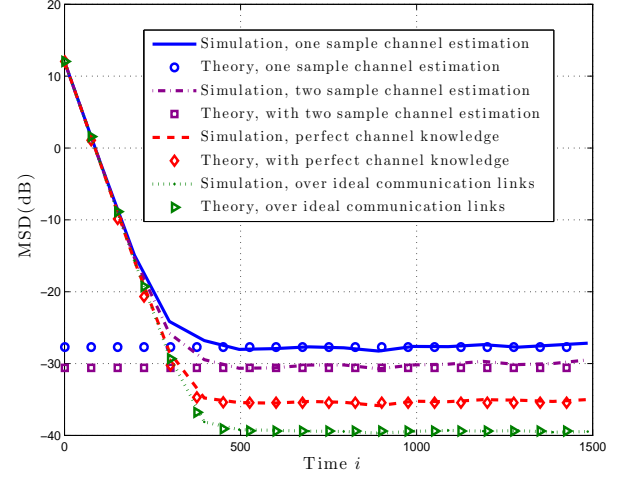


Fig. 5: Learning curves of the network in terms of MSD and EMSE.

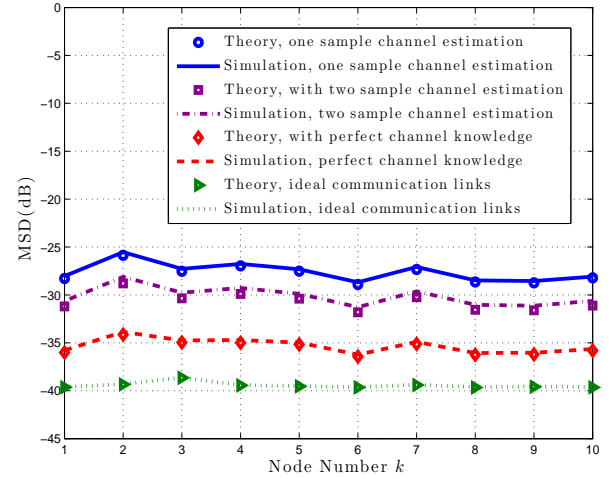


Fig. 6: Steady-state MSD over the network.

not have access to the CSI of their neighbors and obtain this information using one and two samples pilot data. For reference, we also illustrate the performance of ATC diffusion over ideal communication links in which the communication links between nodes are error-free, i.e., for each node  $k$ ,  $\psi_{\ell k, i} = \psi_{\ell, i}$  for all  $i$ .

The best performance in these experiments belongs to the diffusion strategy that runs over network with ideal communication links. As expected, the diffusion strategy with perfect CSI knowledge outperforms diffusion strategy with channel estimation using one or two samples pilot data, respectively, by 5dB and 7dB. In particular, the steady-state mean-square performance of the algorithm improves almost by 2dB for an additional sample of pilot data used for channel estimation. Therefore, if the wireless channels are slowly-varying, by using a larger number of pilot data, it is possible to approach the performance of the diffusion strategy algorithm with

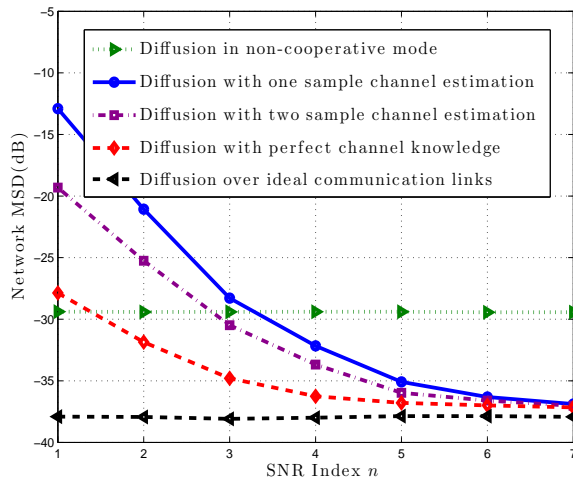


Fig. 7: The network performance comparison with non-cooperative diffusion LMS and with diffusion LMS over ideal communication links.

perfect CSI.

We have also produced a transient MSD curve using standard diffusion LMS [2], under similar fading conditions and noise. The results showed that the network MSD grows unbounded (i.e.,  $\text{error} \rightarrow \infty$ ). This problem can be justified using the fact that some nodes, in the combination step, use severely distorted data from neighbors with bad channel conditions and low SNR. Consequently a large error is introduced into their updated intermediate estimates, which then will propagate into the network in the following iterations and cause catastrophic network failure.

In Fig. 7, we compare the performance of diffusion strategies for different ranges of SNR over the network. We also make some comparisons between the cooperative and non-cooperative networks where in the latter case the network runs a stand-alone LMS filter at each node, which is equivalent to running the diffusion strategy with  $A_i = I$ . In Fig. 7, the SNR index  $n \in \{1, 2, \dots, 7\}$  over the  $x$ -axis refers to the  $n$ -th network SNR distribution, as obtained by uniformly scaling up the initial SNR distribution over the network by 5dB for each increment in the integer  $n$ , as represented by  $\text{SNR}_n = \text{SNR}_{\text{ini}} + 5n$  (dB), where  $\text{SNR}_{\text{ini}}$  are the SNR of the connected nodes illustrated in Fig. 2, and are obtained from uniformly distributed random variables in the range between [5 10]dB.

As shown in Fig. 7, the performance of non-cooperative adaptation and diffusion LMS with ideal communication links remains invariant with changes in the SNR values. This is expected since the performance of the diffusion LMS in these cases is not affected by the communication noise,  $v_{\ell k, i}^{(\psi)}$  and  $v_{\ell k, i}^{(y)}$ . In comparison, the performance of the modified diffusion strategy over wireless links depends on the CSI. As the knowledge about the network CSI increases, the performance im-

proves. From this result, we observe that at low SNR the performance discrepancies between diffusion with perfect CSI and diffusion with channel estimation is larger compared to high SNR scenarios. This difference in performance can be reduced by using more pilot data to estimate the channel coefficients in each time slot. In addition, at very low SNR, we see that the non-cooperative case outperforms the modified diffusion strategy. This result suggests that in wireless networks with high levels of communication noise at all nodes (e.g., when the nodes transmit power is very low), to maintain a satisfactory performance level the network must switch to the non-cooperative mode. This also suggests that if the transmit power of some nodes is below some threshold value, these nodes should go to a sleep mode in order to avoid error propagation over the network.

## 6 CONCLUSION

We extended the application of diffusion LMS strategies to sensor networks with time-varying fading wireless channels. We analyzed the convergence behavior of the modified diffusion LMS algorithms, and established conditions under which the algorithms converge and remain stable in the mean and mean-square error sense. The analysis revealed that the performance of the diffusion strategies highly depend on the level of CSI knowledge and the level of communication noise power over the network. In particular, when the CSI are known, the modified diffusion algorithms are asymptotically unbiased and converge in the slow adaptation regime. In contrast, the parameter estimates will become biased when the CSI are obtained through pilot-aided channel estimation. Nevertheless, the size of the bias can be made small by increasing the number of pilot symbols or increasing the link SNR.

## REFERENCES

- [1] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. on Signal Processing*, vol. 56, no. 7, pp. 3122–3136, July 2008.
- [2] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.
- [3] A. H. Sayed, "Diffusion adaptation over networks," in *E-Reference Signal Processing*, vol. 3, R. Chellappa and S. Theodoridis, Eds., pp. 323–454, Academic Press, 2014., May 2012.
- [4] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May 2013.
- [5] R. Abdolee, B. Champagne, and A. Sayed, "Estimation of space-time varying parameters using a diffusion LMS algorithm," *IEEE Trans. on Signal Processing*, vol. 62, no. 2, pp. 403–418, Jan. 2014.
- [6] R. Abdolee and B. Champagne, "Diffusion LMS algorithms for sensor networks over non-ideal inter-sensor wireless channels," in *Proc. of Int. Conf. on Dist. Computing in Sensor Systems*, Barcelona, Spain, June 2011, pp. 1–6.
- [7] R. Abdolee, B. Champagne, and A. Sayed, "Diffusion LMS for source and estimation in sensor networks," in *IEEE Statistical Signal Processing Workshop (SSP)*, Ann Arbor, Michigan, Aug. 2012, pp. 165–168.

- [8] P. Di Lorenzo and S. Barbarossa, "A bio-inspired swarming algorithm for decentralized access in cognitive radio," *IEEE Trans. on Signal Processing*, vol. 59, pp. 6160–6174, Dec. 2011.
- [9] R. Abdoolee and B. Champagne, "Distributed blind adaptive algorithms based on constant modulus for wireless sensor networks," in *Proc. IEEE Int. Conf. on Wireless and Mobile Commun.*, Sept. 2010, pp. 303–308.
- [10] S. Kar and J. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE J. Selected Topics on Signal Processing*, vol. 5, no. 4, pp. 674–690, Aug. 2011.
- [11] S. Barbarossa and G. Scutari, "Bio-inspired sensor network design," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 26–35, May 2007.
- [12] P. Braca, S. Marano, and V. Matta, "Running consensus in wireless sensor networks," in *Proc. Int. Conf. on Information Fusion*, Cologne, Germany, June 2008, pp. 1–6.
- [13] U. A. Khan and J. M. Moura, "Distributing the Kalman filter for large-scale systems," *IEEE Trans. on Signal Processing*, vol. 56, no. 10, pp. 4919–4935, Oct. 2008.
- [14] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *IEEE Trans. on Signal Process.*, vol. 57, pp. 2365–2382, June 2009.
- [15] T. Aysal, M. Yildiz, A. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Trans. on Signal Processing*, vol. 57, no. 7, pp. 2748–2761, July 2009.
- [16] S. Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. on Signal Processing*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.
- [17] C. G. Lopes and A. H. Sayed, "Diffusion adaptive networks with changing topologies," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process.*, Apr. 2008, pp. 3285–3288.
- [18] L. Li, J. A. Chambers, C. G. Lopes, and A. H. Sayed, "Distributed estimation over an adaptive incremental network based on the affine projection algorithm," *IEEE Trans. on Signal Processing*, vol. 58, no. 1, pp. 151–164, Jan. 2010.
- [19] S. Chouvardas, K. Slavakis, and S. Theodoridis, "Adaptive robust distributed learning in diffusion sensor networks," *IEEE Trans. on Signal Process.*, vol. 59, pp. 4692–4707, Oct. 2011.
- [20] R. Abdoolee, B. Champagne, and A. H. Sayed, "A diffusion LMS strategy for parameter estimation in noisy regressor applications," in *Proc. European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 749–753.
- [21] R. Abdoolee and B. Champagne, "Diffusion LMS strategies in sensor networks with noisy input data," *IEEE/ACM Transaction on Networking*, vol. pp, no. 99, pp. x–x, Sept. 2014.
- [22] R. Abdoolee, B. Champagne, and A. H. Sayed, "Diffusion LMS localization and tracking algorithm for wireless cellular networks," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process.*, Vancouver, Canada, May 2013, pp. 4598–4602.
- [23] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Trans. on Signal Processing*, vol. 60, pp. 4289–4305, Aug. 2012.
- [24] A. Khalili, M. Tinati, A. Rastegarnia, and J. Chambers, "Transient analysis of diffusion least-mean squares adaptive networks with noisy channels," *Int. Jnl. of Adaptive Cont. and Signal Proc.*, Feb. 2012.
- [25] —, "Steady-state analysis of diffusion LMS adaptive networks with noisy links," *IEEE Trans. on Signal Processing*, vol. 60, no. 2, pp. 974–979, Feb. 2012.
- [26] X. Zhao, S. Y. Tu, and A. H. Sayed, "Diffusion adaptation over networks under imperfect information exchange and non-stationary data," *IEEE Trans. on Signal Processing*, vol. 60, no. 7, pp. 3460–3475, July 2012.
- [27] M. R. Gholami, E. G. Strom, and A. H. Sayed, "Distributed estimation over cooperative networks with missing data," in *IEEE Global SIP*, Austin, TX, Dec. 2013, pp. 411–414.
- [28] M. G. Rabbat, R. D. Nowak, and J. A. Bucklew, "Generalized consensus computation in networked systems with erasure links," in *IEEE 6th Workshop on Signal Process. Adv. in Wireless Comm.*, New York, NY, June 2005, pp. 1088–1092.
- [29] D. Jakovetic, J. Xavier, and J. M. Moura, "Weight optimization for consensus algorithms with correlated switching topology," *IEEE Trans. on Signal Process.*, vol. 58, pp. 3788–3801, July 2010.
- [30] K. Chan, A. Swami, Q. Zhao, and A. Scaglione, "Consensus algorithms over fading channels," in *Military Communications Conference*, San Jose, CA, Nov. 2010, pp. 549–554.
- [31] S. Kar and J. M. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. on Signal Processing*, vol. 57, no. 1, pp. 355–369, Jan. 2009.
- [32] —, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Trans. on Signal Processing*, vol. 58, no. 3, pp. 1383–1400, Mar. 2010.
- [33] I. Lobel and A. Ozdaglar, "Distributed subgradient methods for convex optimization over random networks," *IEEE Trans. on Automatic Control*, vol. 56, no. 6, pp. 1291–1306, June 2011.
- [34] A. Rastegarnia, W. M. Bazzi, A. Khalili, and J. A. Chambers, "Diffusion adaptive networks with imperfect communications: link failure and channel noise," *IET Signal Processing*, vol. 8, no. 1, pp. 59–66, 2014.
- [35] R. Abdoolee, B. Champagne, and A. H. Sayed, "Diffusion LMS strategies for parameter estimation over fading wireless channels," in *Proc. of IEEE Int. Conf. on Communications (ICC)*, Budapest, Hungary, June 2013, pp. 1926–1930.
- [36] C. See, R. Abd-Alhameed, Y. Hu, and K. Horoshenkov, "Wireless sensor transmission range measurement within the ground level," in *Antennas and Propagation Conference, 2008. LAPC 2008.*, Loughborough, UK, Sept. 2008, pp. 225–228.
- [37] J. G. Proakis, *Digital Communications*. 4th edition, McGraw-Hill, 2000.
- [38] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*. 2nd edition, Addison-Wesley, 1994.
- [39] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Performance analysis of the consensus-based distributed LMS algorithm," *EURASIP J. Adv. Signal Process.*, pp. 1–19, Nov. 2009.
- [40] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: Link failures and channel noise," *IEEE Trans. on Signal Processing*, vol. 57, no. 1, pp. 355–369, Jan. 2009.
- [41] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.
- [42] N. Takahashi, I. Yamada, and A. H. Sayed, "Diffusion least-mean squares with adaptive combiners: formulation and performance analysis," *IEEE Trans. on Signal Processing*, vol. 58, no. 9, pp. 4795–4810, Sept. 2010.
- [43] A. H. Sayed, *Adaptive Filters*. Wiley-IEEE Press, 2008.
- [44] R. Koning, H. Neudecker, and T. Wansbeek, "Block Kronecker products and the vecb operator," *Linear Algebra and its Applications*, vol. 149, pp. 165–184, Apr. 1991.
- [45] J. Gallier, *Geometric Methods and Applications*. Springer, 2011.



## APPENDIX A

### COMPUTATION OF $R_{v,k}$

To obtain  $R_{v,k}$  in (73), we need to compute the expectation

$$\mathbb{E}[\mathbf{a}_{\ell,k}^2(i) |\hat{\mathbf{g}}_{\ell,k}(i)|^2] = \mathbb{E}\left[\frac{\mathbf{a}_{\ell,k}^2(i)}{\frac{P_t}{r_{\ell,k}^\alpha} |\hat{\mathbf{h}}_{\ell,k}(i)|^2}\right] \quad (108)$$

for  $\ell \in \mathcal{N}_k \setminus k$ . For the case  $\ell = k$ , we have  $R_{v,\ell k} = 0$  and hence the expectation of  $[\mathbf{a}_{\ell,k}^2(i) |\hat{\mathbf{g}}_{\ell,k}(i)|^2] R_{v,\ell k}$  in (73) is zero. For  $\ell \neq k$ , we proceed as follows. Since the joint probability distribution function of the numerator and denominator in (108) is unknown, the expectation can be approximated using one of two ways. In the first method, we can resort to computer simulations. In the second method, we can resort to a Taylor series approximation as follows. We introduce the real-valued auxiliary variable  $\mathbf{x} = \mathbf{a}_{\ell,k}^2(i)$ . Considering the combination rule (22), the expectation of  $\mathbf{x}$  when  $\ell \neq k$  will be:

$$\mathbb{E}[\mathbf{x}] = \gamma_{\ell k}^2 P_{\ell,k} \quad (109)$$

To compute the variance and expectation of the denominator in (108), we let the exponential distribution function  $f_{\mathbf{y}}(y)$  with parameter  $\lambda$  given by (27) denote the pdf of  $\mathbf{y} = |\hat{\mathbf{h}}_{\ell,k}(i)|^2$ , i.e.,

$$f_{\mathbf{y}}(y) = \lambda_{\ell,k} e^{-\lambda_{\ell,k} y}, \text{ for } y \in [0, \infty) \quad (110)$$

We also let  $f_{\mathbf{y}}^{(t)}(y)$  represent the pdf of  $\mathbf{y}$  for  $y \in [\nu_{\ell,k}, \infty)$ . It can be verified that  $f_{\mathbf{y}}^{(t)}(y)$  represents a truncated exponential distribution and is given by:

$$f_{\mathbf{y}}^{(t)}(y) = \lambda_{\ell,k} e^{-\lambda_{\ell,k}(y - \nu_{\ell,k})}, \text{ for } y \in [\nu_{\ell,k}, \infty) \quad (111)$$

If we now define

$$\mathbf{z} = \frac{P_t}{r_{\ell,k}^\alpha} \mathbf{y} \quad (112)$$

Then, the pdf of  $\mathbf{z}$  can be computed as [38]:

$$f_{\mathbf{z}}(z) = \left| \frac{dy}{dz} \right| f_{\mathbf{y}}^{(t)}(g^{-1}(z)) \quad (113)$$

where

$$\frac{dy}{dz} = \frac{r_{\ell,k}^\alpha}{P_t} \text{ and } g^{-1}(z) = \frac{r_{\ell,k}^\alpha}{P_t} z \quad (114)$$

Therefore,

$$f_{\mathbf{z}}(z) = \frac{r_{\ell,k}^\alpha}{P_t} \lambda_{\ell,k} e^{-\lambda_{\ell,k} \left( \frac{r_{\ell,k}^\alpha}{P_t} z - \nu_{\ell,k} \right)}, \text{ for } z \in \left[ \frac{P_t}{r_{\ell,k}^\alpha} \nu_{\ell,k}, \infty \right) \quad (115)$$

Using this distribution the mean and variance of  $\mathbf{z}$  will be [38]:

$$\mathbb{E}[\mathbf{z}] = \frac{P_t}{r_{\ell,k}^\alpha} \left( \frac{1}{\lambda_{\ell,k}} + \nu_{\ell,k} \right) \quad (116)$$

$$\text{var}(\mathbf{z}) = \left( \frac{P_t}{r_{\ell,k}^\alpha \lambda_{\ell,k}} \right)^2 \quad (117)$$

We can now proceed to approximate the expectation (108) by defining

$$f(\mathbf{x}, \mathbf{z}) = \frac{\mathbf{x}}{\mathbf{z}} \quad (118)$$

and employing a second order Taylor series expansion to write:

$$\mathbb{E}[f(\mathbf{x}, \mathbf{z})] \approx \frac{\mathbb{E}[\mathbf{x}]}{\mathbb{E}[\mathbf{z}]} - \frac{1}{(\mathbb{E}[\mathbf{z}])^2} \text{cov}(\mathbf{x}, \mathbf{z}) + \frac{\mathbb{E}[\mathbf{x}]}{(\mathbb{E}[\mathbf{z}])^3} \text{var}(\mathbf{z}) \quad (119)$$

Substituting,  $\mathbb{E}[\mathbf{x}]$ ,  $\mathbb{E}[\mathbf{z}]$ ,  $\text{cov}(\mathbf{x}, \mathbf{z})$  and  $\text{var}(\mathbf{z})$  into (119), we then arrive at:

$$\begin{aligned} \mathbb{E}[f(\mathbf{x}, \mathbf{z})] &\approx \mathbb{E}[\mathbf{a}_{\ell,k}^2(i) |\hat{\mathbf{g}}_{\ell,k}(i)|^2] \\ &\approx \gamma_{\ell k}^2 P_{\ell,k} \left( \frac{1}{\frac{P_t}{r_{\ell,k}^\alpha} \left( \frac{1}{\lambda_{\ell,k}} + \nu_{\ell,k} \right)} - \frac{\nu_{\ell,k}}{\frac{P_t}{r_{\ell,k}^\alpha} \left( \frac{1}{\lambda_{\ell,k}} + \nu_{\ell,k} \right)^2} \right. \\ &\quad \left. + \frac{1}{\frac{P_t}{r_{\ell,k}^\alpha} \lambda_{\ell,k}^2 \left( \frac{1}{\lambda_{\ell,k}} + \nu_{\ell,k} \right)^3} \right) \end{aligned} \quad (120)$$

## APPENDIX B

### DERIVATION OF (80)

First, we note that when  $\mathbf{u}_{k,i}$  are zero mean circular complex-valued Gaussian random vectors and i.i.d. over time, then for any Hermitian matrix  $\Gamma$  of compatible dimensions it holds that [43]:

$$\mathbb{E}[\mathbf{u}_{k,i}^* \mathbf{u}_{k,i} \Gamma \mathbf{u}_{k,i}^* \mathbf{u}_{k,i}] = \beta (R_{u,k} \Gamma R_{u,k}) + R_{u,k} \text{Tr}(\Gamma R_{u,k}) \quad (121)$$

where  $\beta = 1$  for complex regressors and  $\beta = 2$  when the regressors are real. Using (121) and spatial independence of the regression data we have

$$\begin{aligned} \mathbb{E}[\mathbf{u}_{k,i}^* \mathbf{u}_{k,i} \Gamma \mathbf{u}_{\ell,i}^* \mathbf{u}_{\ell,i}] &= R_{u,k} \Gamma R_{u,\ell} + \delta_{k\ell} (\beta - 1) R_{u,k} \Gamma R_{u,k} \\ &\quad + \delta_{k\ell} R_{u,k} \text{Tr}(\Gamma R_{u,k}) \end{aligned} \quad (122)$$

where  $\delta_{k\ell}$  is the Dirac delta sequence. To compute  $\bar{\mathcal{F}}$ , we first introduce

$$\mathcal{L}_i = (I - \mathcal{M} \mathcal{R}_i) \mathcal{Q} (I - \mathcal{M} \mathcal{R}_i) \quad (123)$$

where  $\mathcal{Q}$  is an arbitrary deterministic Hermitian matrix. We now note that

$$\begin{aligned} \text{bvec}(\mathbb{E}[\mathcal{L}_i]) &\stackrel{(i)}{=} \mathbb{E}[(I - \mathcal{M} \mathcal{R}_i)^T \otimes_b (I - \mathcal{M} \mathcal{R}_i)] \text{bvec}(\mathcal{Q}) \\ &\stackrel{(ii)}{=} \bar{\mathcal{F}} \text{bvec}(\mathcal{Q}) \end{aligned} \quad (124)$$

where (ii) obtained by comparing the expectation term on the right hand side of (i) with definition (78). We proceed by taking expectation of both sides of (123), i.e.,

$$\mathbb{E}[\mathcal{L}_i] = \mathcal{Q} - \mathcal{R} \mathcal{M} \mathcal{Q} - \mathcal{Q} \mathcal{M} \mathcal{R} + \mathbb{E}[\mathcal{R}_i \mathcal{M} \mathcal{Q} \mathcal{M} \mathcal{R}_i] \quad (125)$$

To compute the block vectorization of the last term on the right hand side of (125), we introduce the block partitioned matrix  $\mathcal{Q}' = \mathcal{M} \mathcal{Q} \mathcal{M}$  with blocks  $\mathcal{Q}'_{k\ell}$  and use (122) to obtain (126), where  $r_k = \text{vec}(R_{u,k})$ .

Now, using (125), we can write:

$$\begin{aligned} \text{bvec}(\mathbb{E}[\mathcal{L}_i]) &= (I - I \otimes_b \mathcal{M} \mathcal{R} - \mathcal{R}^T \mathcal{M} \otimes_b I) \text{bvec}(\mathcal{Q}) \\ &\quad + \text{bvec}(\mathbb{E}[\mathcal{R}_i \mathcal{Q}' \mathcal{R}_i]) \end{aligned} \quad (127)$$

$$\text{bvec}(\mathbb{E}[\mathcal{R}_i \mathcal{Q}' \mathcal{R}_i]) = \left\{ (\mathcal{R}^T \otimes_b \mathcal{R}) + \sum_{k=1}^N \left[ \text{diag}((\text{vec}(\text{diag}(e_k))) \right] \otimes [(\beta - 1)(R_{k,u}^T \otimes R_{k,u}) + r_k r_k^*] \right\} (\mathcal{M} \otimes_b \mathcal{M}) \text{bvec}(\mathcal{Q}) \quad (126)$$

From (124), (126) and (127) and using the fact that the real vector space of Hermitian matrices is isomorphic to  $\mathbb{R}^{N^2 \times 1}$  [45], we arrive at (80).

## APPENDIX C COMPUTATION OF $\mathcal{D}$

We expand  $\mathcal{D} = \mathbb{E}[\mathcal{D}_i]$  in (79) as:

$$\mathcal{D} = \left\{ \mathbb{E}[\mathbf{A}_i \otimes \mathbf{A}_i] + \mathbb{E}[\mathbf{A}_i \otimes \mathbf{E}_i^{*T}] + \mathbb{E}[\mathbf{E}_i \otimes \mathbf{A}_i] + \mathbb{E}[\mathbf{E}_i \otimes \mathbf{E}_i^{*T}] \right\} \otimes I_{M^2} \quad (128)$$

The  $(r, z)$ -th entry of  $\mathbb{E}[\mathbf{A}_i \otimes \mathbf{A}_i]$ , denoted by  $f_{r,z}$ , is:

$$f_{r,z} = \mathbb{E}[\mathbf{a}_{\ell,k}(i) \mathbf{a}_{m,n}(i)] \quad (129)$$

where the relation between  $(r, z)$  and  $(\ell, k)$  is:

$$r = (\ell - 1)N + m, \text{ and } z = (k - 1)N + n \quad (130)$$

When  $k \neq n$ , entries  $\mathbf{a}_{\ell,k}(i)$  and  $\mathbf{a}_{m,n}(i)$  come from different columns of  $\mathbf{A}_i$  and are independent. Hence, in this case, we can write:

$$f_{r,z} = \mathbb{E}[\mathbf{a}_{\ell,k}(i)] \mathbb{E}[\mathbf{a}_{m,n}(i)] \quad (131)$$

with

$$\mathbb{E}[\mathbf{a}_{j,q}(i)] = \begin{cases} 1 - \sum_{r \in \mathcal{N}_q \setminus q} p_{rq} \gamma_{rq}, & \text{if } j = q \\ p_{jq} \gamma_{jq}, & \text{otherwise} \end{cases} \quad (132)$$

When  $k = n$ , the entries  $\mathbf{a}_{\ell,k}(i)$  and  $\mathbf{a}_{m,n}(i)$  come from the same column of  $\mathbf{A}_i$  and may be dependent. In this case, there are four possibilities:

(1) if  $\ell = m$  and  $\ell \neq k$ :

$$f_{r,z} = \gamma_{\ell,k}^2 p_{\ell,k} \quad (133)$$

(2) if  $\ell = m$  and  $\ell = k$ :

$$f_{r,z} = \mathbb{E} \left[ \left( 1 - \sum_{\ell \in \mathcal{N}_k \setminus k} \mathbf{a}_{\ell,k}(i) \right) \left( 1 - \sum_{\ell \in \mathcal{N}_k \setminus k} \mathbf{a}_{\ell,k}(i) \right) \right] \quad (134)$$

$$= 1 - 2 \sum_{\ell \in \mathcal{N}_k \setminus k} p_{\ell,k} (\gamma_{\ell,k} - \gamma_{\ell,k}^2) - \sum_{\ell \in \mathcal{N}_k \setminus k} p_{\ell,k}^2 \gamma_{\ell,k}^2 + \sum_{(\ell \in \mathcal{N}_k \setminus k) (m \in \mathcal{N}_k \setminus k)} p_{\ell,k} p_{m,k} \gamma_{\ell,k} \gamma_{m,k} \quad (135)$$

(3) if  $\ell \neq m$  and  $\ell \neq k$  and  $m \neq n$ :

$$f_{r,z} = \gamma_{\ell,k} \gamma_{m,n} p_{\ell,k} p_{m,n} \quad (136)$$

(4) if  $\ell \neq m$  and  $\ell = k$  and  $m \neq n$ :

$$f_{r,z} = \mathbb{E} \left[ \left( 1 - \sum_{j \in \mathcal{N}_k \setminus k} \mathbf{a}_{j,k}(i) \right) \mathbf{a}_{m,n}(i) \right] = \gamma_{m,n} p_{m,n} \left( 1 - \gamma_{m,n} + \sum_{j \in \mathcal{N}_k \setminus \{k,m\}} \gamma_{j,k} p_{j,k} \right) \quad (137)$$

The  $(r, z)$ -th entry of  $\mathbb{E}[\mathbf{A}_i \otimes \mathbf{E}_i^{*T}]$ , denoted by  $x_{r,z}$ , can be expressed as:

$$\begin{aligned} x_{r,z} &= -\mathbb{E} \left[ \mathbf{a}_{\ell,k}(i) \mathbf{a}_{m,n}(i) \hat{\mathbf{g}}_{m,n}^*(i) \mathbf{v}_{m,n}^{(y)*}(i) \right] \\ &= -\mathbb{E} \left[ \mathbf{a}_{\ell,k}(i) \mathbf{a}_{m,n}(i) \frac{\sqrt{\frac{r^\alpha}{P_t}} \mathbf{h}_{m,n}(i) \mathbf{v}_{m,n}^{(y)*}(i) + \frac{r^\alpha}{P_t} |\mathbf{v}_{m,n}^{(y)}(i)|^2}{|\mathbf{h}_{m,n}(i) + \sqrt{\frac{r^\alpha}{P_t}} \mathbf{v}_{m,n}^{(y)}(i)|^2} \right] \\ &\quad \left[ |\mathbf{h}_{m,n}(i) + \sqrt{\frac{r^\alpha}{P_t}} \mathbf{v}_{m,n}^{(y)}(i)|^2 \geq \nu_{m,n} \right] \end{aligned} \quad (138)$$

Likewise, the entries of  $\mathbb{E}[\mathbf{E}_i \otimes \mathbf{A}_i]$  and  $\mathbb{E}[\mathbf{E}_i \otimes \mathbf{E}_i^{*T}]$  can be expressed in terms of the combination weights, channel coefficients and the estimation error. We can follow the argument presented in Remark 3 to show that the right hand side of (138) as well as the entries of  $\mathbb{E}[\mathbf{E}_i \otimes \mathbf{A}_i]$  and  $\mathbb{E}[\mathbf{E}_i \otimes \mathbf{E}_i^{*T}]$  are invariant with respect to time and have finite values.